Anastasia Lavrenko

Compressive Acquisition and Processing of Sparse Analog Signals

Anastasia Lavrenko

Compressive Acquisition and Processing of Sparse Analog Signals

Universitätsverlag Ilmenau
2018

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über http://dnb.d-nb.de abrufbar.

Diese Arbeit hat der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität Ilmenau als Dissertation vorgelegen.

Tag der Einreichung:	7. November 2017
1. Gutachter:	Univ.-Prof. Dr.-Ing. habil. Reiner S Thomä (Technische Universität Ilmenau)
2. Gutachter:	Prof. Dr.-Ing. Marius Pesavento (Technische Universität Darmstadt)
3. Gutachter:	Prof. Dr. Orhan Arikan (Bilkent Universität, Ankara, Türkei)
Tag der Verteidigung:	2. Mai 2018

Technische Universität Ilmenau/Universitätsbibliothek
Universitätsverlag Ilmenau
Postfach 10 05 65
98684 Ilmenau
http://www.tu-ilmenau.de/universitaetsverlag

readbox unipress
in der readbox publishing GmbH
Am Hawerkamp 31
48155 Münster
http://unipress.readbox.net

ISBN 978-3-86360-183-6 (Druckausgabe)
URN urn:nbn:de:gbv:ilm1-2018000167

Titelfoto: photocase.com | AlexFlint

Abstract

Since the advent of the first digital processing units, the importance of digital signal processing has been steadily rising. Today, most signal processing happens in the digital domain, requiring that analog signals be first sampled and digitized before any relevant data can be extracted from them. For decades, conventional uniform sampling that is governed by the Nyquist sampling theorem has provided an almost universal means to this end. The recent explosion of the demands for data acquisition, storage and processing, however, has pushed the capabilities of conventional acquisition systems to their limits in many application areas. By offering an alternative view on the signal acquisition process, ideas from sparse signal processing and one of its main beneficiaries compressed sensing (CS), have the potential to assist alleviating some of these problems. Building on the premise that the signal information rate is often much lower than what is dictated by its native representation, CS provides an alternative acquisition and processing framework that attempts to reduce the sampling rate while preserving the information content of the signal.

In this thesis, we explore some of the basic foundations of the finite-dimensional CS framework and its connection to sub-Nyquist sampling and processing of sparse continuous analog signals with application to multiband sensing. Despite being a focus of active research for over a decade, there still remain significant gaps in understanding the implications that compressive approaches have on the signal recovery and processing performance, especially against noisy settings and in relation to practical sampling problems. This dissertation aims at filling some of these gaps. More specifically, we look into the ways the application of a compressive measurement kernel impacts signal and noise characteristics and the relation it has to the signal recovery performance. We also investigate methods to infer the current complexity of the signal scene from the reduced-rate compressive observations without resorting to Nyquist-rate processing and show the advantage this knowledge offers to the recovery process. Having considered some of the universal aspects of compressive systems, we then move to studying a particular application, namely that of sub-Nyquist sampling and processing of sparse analog multiband signals. Within the sub-Nyquist sampling framework, we examine three different multiband scenarios that involve multiband sensing in spectral, angular and spatial domains. For each of them, we provide a sub-Nyquist receiver architecture, develop recovery methods and numerically evaluate their performance.

Zusammenfassung

Seit dem Aufkommen der ersten digitalen Verarbeitungseinheiten hat die Bedeutung der digitalen Signalverarbeitung stetig zugenommen. Heutzutage findet die meiste Signalverarbeitung im digitalen Bereich statt, was erfordert, dass analoge Signale zuerst abgetastet und digitalisiert werden, bevor relevante Daten daraus extrahiert werden können. Jahrzehntelang hat die herkömmliche äquidistante Abtastung, die durch das Nyquist-Abtasttheorem bestimmt wird, zu diesem Zweck ein nahezu universelles Mittel bereitgestellt. Der kürzliche explosive Anstieg der Anforderungen an die Datenerfassung, -speicherung und -verarbeitung hat jedoch die Fähigkeiten herkömmlicher Erfassungssysteme in vielen Anwendungsbereichen an ihre Grenzen gebracht. Durch eine alternative Sichtweise auf den Signalerfassungsprozess können Ideen aus der sparse Signalverarbeitung und einer ihrer Hauptanwendungsgebiete, Compressed Sensing (CS), dazu beitragen, einige dieser Probleme zu mindern. Basierend auf der Annahme, dass der Informationsgehalt eines Signals oft viel geringer ist als was von der nativen Repräsentation vorgegeben, stellt CS ein alternatives Konzept für die Erfassung und Verarbeitung bereit, das versucht, die Abtastrate unter Beibehaltung des Signalinformationsgehalts zu reduzieren.

In dieser Arbeit untersuchen wir einige der Grundlagen des endlichdimensionalen CS-Frameworks und seine Verbindung mit Sub-Nyquist Abtastung und Verarbeitung von sparsen analogen Signalen. Obwohl es seit mehr als einem Jahrzehnt ein Schwerpunkt aktiver Forschung ist, gibt es noch erhebliche Lücken beim Verständnis der Auswirkungen von komprimierenden Ansätzen auf die Signalwiedergewinnung und die Verarbeitungsleistung, insbesondere bei rauschbehafteten Umgebungenen und in Bezug auf praktische Messaufgaben. In dieser Dissertation untersuchen wir, wie sich die Anwendung eines komprimierenden Messkerns auf die Signal- und Rauschcharakteristiken auf die Signalrückgewinnungsleistung auswirkt. Wir erforschen auch Methoden, um die aktuelle Signal-Sparsity-Order aus den komprimierten Messungen abzuleiten, ohne auf die Nyquist-Raten-Verarbeitung zurückzugreifen, und zeigen den Vorteil, den sie für den Wiederherstellungsprozess bietet. Nachdem gehen wir zu einer speziellen Anwendung, nämlich der Sub-Nyquist-Abtastung und Verarbeitung von sparsen analogen Multibandsignalen. Innerhalb des Sub-Nyquist-Abtastung untersuchen wir drei verschiedene Multiband-Szenarien, die Multiband-Sensing in der spektralen, Winkel und räumlichen-Domäne einbeziehen.

Acknowledgements

I would like to thank my supervisor prof. Reiner Thomä for providing me with an opportunity to pursue my doctorate studies at the department of Electronic Measurements at TU Ilmenau, as well as his support and trust in me throughout the years. I am also deeply thankful to my co-supervisor prof. Giovanni Del Galdo for introducing me to the research area that has eventually become the focus of this thesis, and for providing guidance and encouragement along the way. Dr. Florian Römer has my eternal gratitude for his genuine scientific curiosity, originality of thought, professional competence and having a unique ability to make research fun. Without his knowledge and support, this thesis would hardly have been possible.

I would also like to express my appreciation to all the colleagues I was lucky enough to collaborate with over the years, both inside and outside of the institute. In no particular order, thanks to Rajesh Kumar Sharma, Mohamed Ibrahim, Christopher Schirmer, Jonas König, Paulo Dos Santos, and Oleksii Skoblikov from TU Ilmenau, Andreas Bollig and Johannes Schmitz from RWTH Aachen, prof. Orhan Arikan from Bilkent University in Ankara Turkey, prof. Yonina C Eldar, Shahar Stein-Ioushua and Debora Cohen from Technion University of Technology in Haifa Israel, Cristina Gomez and prof. Roberto Hincapi from UPB Medellin Colombia, and many more. I am especially grateful to Christian Schneider, Paolo Garcia and Mariana Pralon for providing a friendly and comfortable working atmosphere at our shared office at TU Ilmenau, as well as the members of our "Mittagspause" group Estefania Cano, Snezhana Jovanoska and Rudolf Zetic for bringing diversity and entertainment into my every day working life, and of course, for being my dear friends.

Last but not least, I thank my family and friends. You have enriched my life and made this journey so much more enjoyable. My final thanks go to my partner Jody Hatton for his unconditional support and unbreakable faith in me. Thank you!

Ilmenau, 2017

Anastasia Lavrenko

to my grandfather Y. Lavrenko.

Inhaltsverzeichnis

Chapter 1

Introduction

1.1 Motivation

The past several decades have witnessed an unprecedented expansion of the *digital world*. Signal digitization (sampling and quantization) and digital domain signal processing have become key blocks of most modern devices, technological processes and even services. The digital world has come so far as to change irreversibly the way we communicate with each other and the world around us. At the heart of this digital revolution lie the theoretical works of Kotelnikov, Nyquist, Shannon and Whittaker that have culminated in the celebrated Shannon-Nyquist sampling theorem. It dictates the minimum sampling rate required for the lossless sampling and recovery of bandlimited analog signals including by extension images, video, and other data types. For almost half a century, signal acquisition in digital sensing and processing systems has been almost universally governed by the principles of (uniform) Nyquist sampling. This has enabled many of the remarkable achievements in today's signal processing systems, ranging from communication devices to digital cameras to robotics and biomedical systems. Despite these achievements, the explosive growth of digital applications presents a number of serious challenges, particularly with respect to required sampling rates, data storage and signal processing capabilities. Altogether, this creates a need for more effective approaches to signal acquisition that are better tailored to particular applications, e.g., by taking into account the knowledge of the underlying signal structure.

Among the different views on possible solutions to the data avalanche problem that signal processing systems are facing nowadays, one stands out: compressed sensing (CS), the sampling paradigm that unites the processes of signal acquisition and signal compression. Originally developed in the context of solving systems of linear equations, CS builds on ideas from nonlinear approximation [1], sparse representations [2]–[7] and lossy compression [8], [9]. In contrast to traditional views on the data compression that perceives it as a task of analyzing an acquired signal and approximating it with as few components as possible, CS provides a different outlook by raising a thought-provoking question: if what one cares for in a signal is only a few information encoding coefficients, could one use this knowledge

to acquire it more efficiently (e.g. at a lower rate) and still be able to extract them? The answer provided by the seminal works of Donoho, Candés, Baraniuk and Tao imagines a new a digital signal acquisition system: one that combines digitization and compression in a single organic process rather than separating them into two autonomous ones [10]–[15]. The potential of the CS framework was first recognized in image processing applications [16]–[18], but it soon became clear that it has much wider ramifications. Of course, the ideas that lie at its heart are not entirely new. It has long been recognized that alternative approaches to signal sampling can be beneficial in certain scenarios [19]–[22]. Furthermore, since the works of Henri Landau in the late 1960s, there has been an understanding that the minimal sampling rate required for exact recovery of a bandlimited signal is defined by its spectrum occupancy rather than by the overall bandwidth. Landau's theoretical works laid the foundation for the theory of non-uniform sampling and paved the way to the more recent results of Bresler, Feng, and Venkataramani on reduced rate non-uniform sampling [22], [23] which were eventually related to the CS paradigm.

This being said, one of the biggest achievements of CS (besides having inspired an impressive array of theoretical results and numerical tools) is that it creates a unified framework for the recovery of structured high-dimensional signals from their low-dimensional observations. On the other hand, being a mathematical framework that operates mostly within a finite-dimensional world, it generally gives little insight into how such a compressive sampling process can be realized in practice, having in mind acquisition of analog continuous signals. This seeming discrepancy has recently attracted significant research whose results we broadly refer to as the theory of sub-Nyquist sampling[1], which is concerned with sampling continuous signals that lie in the union of subspaces [24]. This dissertation takes further steps towards this goal by expanding the understanding of the CS acquisition process and its implications to practical sampling problems. It explores the foundations of both the finite-dimensional CS world and the sub-Nyquist world of the union of subspaces, their differences and interconnections, on a particular example of compressive acquisition and processing of wideband analog signals with multiband structure. It is worth noting that efficient sampling and recovery of multiband signals has been an area of active research since the works of Landau. Over the years, it not only retained relevance but also multiplied the range of applications it can potentially serve. Thus, nowadays multiband sensing is central to cognitive radio (CR), spectrum monitoring and radio surveillance. It also finds its place in tactical, maritime and underwater communications [25]–[27], as well as in radar [28],

[1] We would like to point out that there is no universal consensus among researchers about the terminology here. Thus, many do not clearly distinguish sub-Nyquist sampling but rather place it under the umbrella of CS. In contrast, in this work we treat CS as a finite-dimensional framework while using the term "sub-Nyquist sampling" in reference to the methods that explicitly consider the acquisition and processing of continuous signals at the rates below Nyquist. The latter, of course, includes the aforementioned earlier works on non-uniform sampling that were not originally developed within the CS framework. Nevertheless, the areas of CS and sub-Nyquist sampling are closely connected and, as such, they benefit and borrow from each others' developments. Ultimately, their successful combination shall one day provide a comprehensive view on the compressive acquisition of analog signals that have sparse or compressible representations.

[29] and ultra wideband (UWB) systems [30]–[33]. With this in mind, this thesis studies the application of the compressive acquisition paradigm to multiband sensing in spectral, angular and spatial domains. Before moving to particular sensing scenarios however, it first examines some of the basic principles that govern the operation of a compressive sensing system providing novel insights into the relations between its main components that are relevant to a wide range of CS applications.

1.2 Thesis outline

This thesis is conceptually organized into two parts that are further arranged into individual chapters. The first part, entitled "CS System Design and Analysis," consists of three chapters that investigate the main blocks of the generic compressive acquisition system and their influence on recovery performance. The first of them, Chapter 2 "Compressed sensing fundamentals," contains a brief summary of the main components constituting the CS framework followed by a discussion of a connection between the classical finite-dimensional CS formulation and the sub-Nyquist sampling of analog signals. The organization of this overview chapter is largely inspired by the book of Prof. Yonina C. Eldar on compressed sensing and its applications [34]. It is followed by Chapter 3, "Sensing matrix analysis and design," which presents a study of the influence of the compressive measurement kernel on system characteristics such as the effective signal-to-noise ratio (SNR) at the output of the (compressive) receiver and the impact it has on recovery performance. The chapter proceeds by exploring possible methods of measurement design optimization in order to obtain a kernel suitable for practical implementation. Chapter 4, "Sparsity order estimation," completes the first part by discussing the task of sparsity order estimation from compressed observations. The sparsity of the input signal is of vital importance to the CS system, while its knowledge can be used to assist the recovery process or adapt the measurement strategy. This said, the order estimation task in this thesis is considered in both static and dynamic scenarios, and it is accompanied by a numerical demonstration of the performance improvement it offers.

The second part of this thesis, "Applications in Sub-Nyquist Sampling," is dedicated to an application of compressive acquisition ideas to sub-Nyquist sampling and processing of sparse analog multiband signals. It begins with the introduction of a multiband signal model and an associated wideband spectrum sensing (WSS) problem in Chapter 5, "Wideband multiband spectrum sensing," followed by a brief review of WSS approaches. It then moves to considering a particular architecture for sub-Nyquist sampling of analog signals, the so-called modulated wideband converter (MWC), and formulating the WSS task in the CS framework. It demonstrates that the spectral occupancy can be coarsely estimated from the support of the input signal without the need for full signal/spectrum reconstruction. Chapter 6, "Multiband spectrum sensing and DoA estimation," extends the concept of multiband sensing into the angular domain by adding an array of antennas at the receiver and utilizing the signal structure in both frequency and space. Building on the MWC architecture, this chapter

develops a receiver system for joint frequency-angular sensing and provides an analysis of the conditions for signal recovery and a discussion of possible recovery methods. Finally, Chapter 7, "Multiband spectrum sensing and time delay estimation," considers multiband sensing in a distributed setup with multiple sensing nodes that are time-synchronized with each other. It presents an approach for joint multiband sensing and time delay estimation from sub-Nyquist samples and shows how the distributed multiband measurements can help in determining in-band channel interference. In conclusion, Chapter 8, "Conclusions and outlook," summarizes the main thesis findings and highlights related research questions that are open to further investigation.

1.3 Summary of contributions

On a conceptual level, this thesis contributes to a better understanding of the connections between the components of the finite-dimensional CS framework, their relation and influence on the recovery performance, and the elements of practical systems for sub-Nyquist sampling of continuous sparse signals. Each chapter examines specific aspects of the design of a compressive acquisition system, while exploring their interconnections and implications on the operation of the whole system. Major contributions of each chapter are listed below.

1. **Chapter 3 "Sensing matrix analysis and design".** The first contribution of this chapter is the uncovering of new aspects in the relation between the choice of the sensing matrix in CS and the recovery performance. By taking a deterministic view on the composition of the sensing matrix, it provides novel insights into the fundamental differences between compressive and traditional (Nyquist-rate) signal acquisition. This leads to the second major contribution of this chapter, which is the introduction of an alternative approach to sensing matrix design. Instead of relying on random ensembles, we propose optimizing the sensing matrix with respect to particular application requirements at hand and provide examples of possible design formulations.

2. **Chapter 4 "Sparsity order estimation".** The main contribution of this chapter is the development of an eigenvalue-based theory of sparsity order estimation from compressive observations. It encompasses two most common CS signal models, the so-called multiple measurement vector (MMV) and the single measurement vector (SMV), and can be applied in dynamic scenarios. Thus, in this chapter we investigate the sparsity order problem in detail, propose particular estimation methods for each considered scenario and evaluate the potential of order estimation for recovery assistance.

3. **Chapter 5 "Wideband multiband spectrum sensing".** Chapter 5 connects the finite-dimensional CS theory with the sub-Nyquist sampling framework on the example of wideband acquisition and processing of analog multiband signals. It presents a

compressive energy detector for coarse estimation of the spectral occupancy from sub-Nyquist samples, contributes to the analysis of the sensing performance in a multiband setting, and relates the findings from previous chapters to a particular application.

4. **Chapter 6 "Multiband spectrum sensing and DoA estimation".** The main contribution of Chapter 6 is the development of a sub-Nyquist system for multiband sensing with antenna arrays. Along with proposing the receiver structure, we formulate the recovery conditions and devise particular reconstruction methods for joint angular-frequency sensing. The proposed methods are largely independent of the structure of the antenna responses and can be applied to practical (e.g., measured) antenna arrays.

5. **Chapter 7 "Multiband spectrum sensing and time delay estimation".** Finally, in Chapter 7 we extend our analysis to a distributed scenario and develop methods for multiband time delay estimation from sub-Nyquist samples provided by multiple time- synchronized sensors. The proposed methods allow recovering relative autocorrelation functions of the signals within the individual sub-bands constituting the multiband input, which is a prerequisite for later time delay estimation.

The contributions discussed above are partially presented in the following peer-reviewed journal [A1]–[A4] and conference [A5]–[A20] publications. The relation of the individual publications to the particular parts of the thesis are provided at the beginning of the respective chapters.

1.4 Notations

We complete this chapter by introducing some common notations used throughout the thesis.

Scalars

x, y, z	scalars
$x(t)$	a continuous function of some continuous variable t
$x[n]$	a discrete function of some discrete variable n
$\mathbb{E}\{x\}$	the expected value of x
$\mathrm{var}\{x\}$	the variance of x
$\mathrm{Re}\{x\}$, $\mathrm{Im}\{x\}$	the real and imaginary parts of x

Vectors and matrices

$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$	vectors
x_i	the ith element of $\boldsymbol{x}$
$\boldsymbol{x}_{\mathcal{I}}$	a sub-vector of $\boldsymbol{x}$ formed by taking the elements indexed by $\mathcal{I}$
$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$	matrices
$\boldsymbol{X}_{i,j} = [\boldsymbol{X}]_{(i,j)}$	an (i, j) element of $\boldsymbol{X}$
$[\boldsymbol{X}]_{n:k}$	a sub-matrix of $\boldsymbol{X}$ formed by taking its nth to kth rows
$\boldsymbol{X}_{\mathcal{I}}$	a sub-matrix of $\boldsymbol{X}$ formed by taking the columns indexed by $\mathcal{I}$
$\boldsymbol{x}^*$, $\boldsymbol{X}^*$	the (element-wise) complex conjugate of $\boldsymbol{x}$, $\boldsymbol{X}$
$\lvert\boldsymbol{x}\rvert$, $\lvert\boldsymbol{X}\rvert$	the (element-wise) absolute value of $\boldsymbol{x}$, $\boldsymbol{X}$
$\mathbb{E}\{\boldsymbol{x}\}$, $\mathbb{E}\{\boldsymbol{X}\}$	the (element-wise) expected value of $\boldsymbol{x}$, $\boldsymbol{X}$
$\mathrm{Re}\{\boldsymbol{x}\}$, $\mathrm{Re}\{\boldsymbol{X}\}$	the (element-wise) real part of $\boldsymbol{x}$, $\boldsymbol{X}$
$\mathrm{Im}\{\boldsymbol{x}\}$, $\mathrm{Im}\{\boldsymbol{X}\}$	the (element-wise) imaginary part of $\boldsymbol{x}$, $\boldsymbol{X}$
$\boldsymbol{x}^{\mathrm{T}}$, $\boldsymbol{X}^{\mathrm{T}}$	a transpose of $\boldsymbol{x}$, $\boldsymbol{X}$
$\boldsymbol{x}^{\mathrm{H}}$, $\boldsymbol{X}^{\mathrm{H}}$	a Hermitian (conjugate) transpose of $\boldsymbol{x}$, $\boldsymbol{X}$
$\boldsymbol{X}^{\dagger} = (\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{X}^*$	a Moore-Penrose matrix pseudo-inverse
$\mathbf{1}_N$, $\mathbf{0}_N$	an all-one, all-zero vector of length N
$\mathbf{I}_N$	an $N \times N$ identity matrix
$\otimes$, $\odot$, $\diamond$	the Kronecker, Hadamard (element-wise) and Khatri-Rao products

Further details on vector and matrix operations are provided in Appendix C.

Tensors and sets

$\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{Z}}$	tensors
$\boldsymbol{\mathcal{X}}_{p \times n \times m}$	a tensor of size $p \times n \times m$
$\boldsymbol{\mathcal{I}}_{3 \times p}$	an identity tensor of size $p \times p \times p$
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	generic sets
$\mathbb{X}$	a special (number) set
$\mathbb{X}^{N \times M}$	a special (number) set of dimensions $N \times M$
$\mathbb{N}$	the set of natural numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{Z}$	the set of integer numbers

Generic sets are defined using curly brackets by either specifying the conditions of belonging to the set, as in $\mathcal{X} = \{x \in \mathbb{R} : x \geq 0\}$, or listing their elements explicitly, e.g., $\mathcal{X} = \left\{\frac{1}{3}, \frac{2}{3}, \ldots, \frac{N}{3}\right\}$ or $\mathcal{X} = \left\{\frac{k}{3}\right\}_{k=1}^{N}$.

Part I

CS System Design and Analysis

In the first part of this thesis, we explore some basic elements of compressed sensing (CS) framework that are largely independent of the application. While being often criticized for regarding the signal model and the measurement process solely from a finite-dimensional perspective, the CS framework provides a variety of tools that, applied in a sensible way, can be employed for solving more practical (infinite-dimensional) problems. In addition to a significant amount of results available to build on, studying a finite-dimensional CS setting offers an advantage of a relatively straightforward, albeit perhaps simplified, way to interpret them. The following chapters, however, do not only serve as an introduction to the finite-dimensional CS world. They also provide novel insights into some fundamental aspects of the design and operation of the CS systems that are essential for practical implementation. One of these is concerned with the influence of the choice of the measurement kernel on the system's parameters and, subsequently, on the recovery performance. Another one is the ability to infer the signal complexity from the compressed observations and to use this knowledge to assist the recovery. Being expressed against a generic CS setting both of them have wide ramifications for a variety of practical applications, one of which, namely that of the recovery and processing of sparse continuous multiband signals, becomes the focus of the second part of this work.

Chapter 2

Compressed sensing fundamentals

2.1 Where analog meets discrete

For decades, there has been an almost universal trend in signal processing that can be roughly summarized as "first sample, sort out later". It is characterized by the transfer of most of the signal processing and analysis to the digital domain coupled to the ever-growing demands to the sampling capabilities of sensing systems. Traditional uniform sampling ensures loseless discrete representation of analog signals, provided that the samples are taken at twice the highest frequency they contain, and provides a straightforward relation between signal properties and operations in analog and digital domains. Thus, it comes as no surprise that uniform Nyquist-rate sampling has become a universal sampling strategy for most of the modern sensing and processing systems. It has been long recognized however that in many practical applications what one is often interested in is some (limited) number of coefficients or signal parameters rather than the exact representation of the signal that they are encoded in. Hence, after the signal is sampled it usually undergoes a process of signal analysis and compression with the goal of extracting and retaining only relevant content. When this is much lower than what is allowed by the total measure of the parameter space defining the signal, we say that the signal can have a more compact (sparse) representation than the one is provided by the traditional Nyquist sampling.

Mathematically, we can express this by introducing an operator $\Psi : \mathcal{D} \to \mathcal{S}$ that maps some parameter (coefficient) space $\mathcal{D} \subseteq \mathbb{R}^N$ to a signal space $\mathcal{S} \subset L_2$ such that

$$\forall\, s(t) \in \mathcal{S}\; s(t) = \Psi(\boldsymbol{d}) : \boldsymbol{d} \in \mathcal{D}. \tag{2.1}$$

Here, $\boldsymbol{d}$ is the coefficient vector that defines the analog signal $s(t)$ in the parameter space $\mathcal{D}$. Depending on N, the dimension of the signal space can be significantly large as it has to provide enough degrees of freedom to account for all possible parameter choices [34]. When the signal exploits only a small number of available degrees of freedom we can represent it as

belonging to some $\mathcal{S}_\gamma \subset \mathcal{S}$ that is characterized by a (lower-dimensional) subspace $\mathcal{D}_k \subseteq \mathbb{R}^K$

$$s(t) \in \mathcal{S} \iff \exists \gamma_0 \in \Gamma \; s(t) \in \mathcal{S}_{\gamma_0} \text{ and } \boldsymbol{d} \in \mathcal{D}_k, \tag{2.2}$$

where Γ denotes an index set and K is assumed to be smaller than N. The model (2.2) is known as the *union of subspaces* since, according to it, the signal $s(t)$ resides in a subspace of $\mathcal{S}$ formed by the union over all $\mathcal{S}_\gamma$, i.e, $s(t) \in \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma$ [24], [35]–[39]. If the exact subspace index γ_0 is known we can apply conventional sampling guided by the dimensions of $\mathcal{S}_{\gamma_0}$ and $\mathcal{D}_k$. When this is not the case, we can still use the traditional framework but with a sampling rate dictated by the dimension of the entire signal space $\mathcal{S}$ instead. Intuitively, this would be a wasteful process as we know that $s(t)$ does not occupy the entire space. Until recently, however, such an approach (sampling the signal with respect to the entire signal subspace followed by identifying its lower-dimensional structure later during the post-processing step) was an almost universal way of treating signals of form (2.2) when γ_0 is unknown.

Motivated by this disregard of the signal structure and subsequent seemingly inefficient signal acquisition step, an alternative framework of compressed sensing (CS) has emerged. It attempts to combine the processes of signal measurement and compression to allow for lower sampling rates that are driven by the amount of active signal information content. To describe this process, we can introduce a linear sampling operator $\tilde{\Phi}$ that maps the (continuous) functional $s(t)$ to an M-element vector of (discrete) measurements $\boldsymbol{y}$

$$\boldsymbol{y} = \tilde{\Phi}\left(s(t)\right). \tag{2.3}$$

At the later step of signal reconstruction, the prior knowledge of the signal low-dimensional structure (e.g., in the form of (2.2)) is utilized to recover $s(t)$ or the information within. The difference between the two approaches is schematically illustrated in Figure 2.1 with the traditional signal acquisition and processing flow sketched in Figure 2.1a and its compressive counterpart depicted in Figure 2.1b.

It is important to note that the complexity reduction at the sensor in terms of the required sampling rates and the amount of generated data promised by CS is paid for by the increase of the computational power needed at the processing phase due to highly non-linear recovery methods. This trade-off between the sensing and processing complexity sets boundaries on the practical applications of CS-based systems. Another restriction comes from the fact that CS, as a mathematical framework, has initially grown from the finite-dimensional algebra which is not always directly applicable to continuous analog signals. If formulated carefully however[1], a discrete model in many cases provides a functional description of the analog sampling system. The main advantage of analyzing the discrete model is that it often allows immediate application of the obtained results to a practical system, with only some mild

[1]We will see how a discrete CS formulation can be used to describe practical sub-Nyquist sampling systems in more detail in Section 2.4 of this chapter as well as in the second part of this thesis.

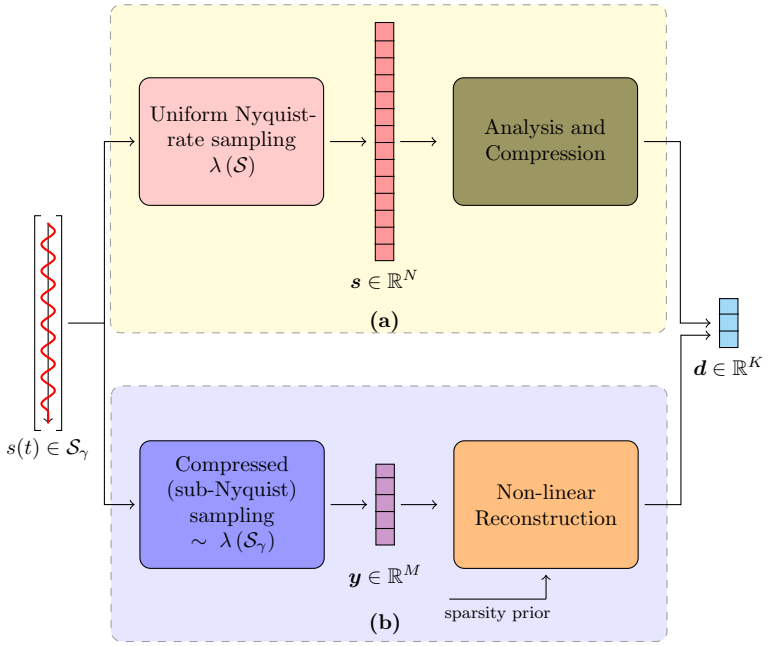


Figure 2.1: Traditional (a) and compressive (b) approaches to signal acquisition and compression. The notation $\lambda(\cdot)$ is used to indicate the Lebesgue measure of a set.

modifications required. For these reasons, it is common to study sub-Nyquist sampling systems in a purely discrete setting. To do so, one assumes that the signal $s(t)$ is initially regularly sampled at the Nyquist rate and then compressed by a linear operator $\boldsymbol{\Phi}$ that performs a dimensionality reduction (from N, that is needed for Nyquist-rate sampling of $s(t)$, to M). Following this approach, we arrive at the classical discrete CS model

$$\boldsymbol{y} = \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{s}, \tag{2.4}$$

where $\boldsymbol{s}$ is a length N vector that contains Nyquist-rate samples of $s(t)$ and $\boldsymbol{\Phi}$ is a measurement matrix of dimensions $N \times M$. Importantly, an additional requirement to $\boldsymbol{s}$ is that it has a sparse representation, i.e., that the number of degrees of freedom it enjoys is small compared to its ambient dimensionality.

In the remainder of this chapter, we briefly review some of the fundamental results in CS that will provide a necessary background for the following chapters. We start the overview with the classical finite-dimensional framework of (2.4) and complete it by examining its extensions to sub-Nyquist sampling of continuous analog signals. Note that we do not aim at providing a comprehensive overview of the CS theory in this chapter, but rather highlight its main elements essential for further discussion. For more detailed analysis we refer the reader to [24], [34], [40] and references therein.

2.2 Single measurement vector

The main goal associated with (2.4) is the recovery of $\boldsymbol{s}$ from its measurements $\boldsymbol{y}$. However, since the system (2.4) has more unknowns (N) than equations (M), it is underdetermined and generally has an infinite number of solutions; this means that for any input signal $\boldsymbol{s}_0$ there exists an infinite amount of signals $\boldsymbol{s}_k$ that result in the same measurement vector $\boldsymbol{y}_0 = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{s}_0 = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{s}_k, s_0 \neq \boldsymbol{s}_k$. What makes having a unique solution possible is the low-dimensional structure of $\boldsymbol{s}$ introduced in the previous section. In CS, this structure is usually captured by the notion of sparsity.

2.2.1 Sparsity models

Sparse and compressible signals. The simplest and most widespread sparsity model is that of strictly sparse signals that can be represented (or well-approximated) by a small number of non-zero coefficients.

Definition 1. *We say that some N-dimensional signal $\boldsymbol{s} \in \mathbb{R}^N$ is K-sparse if exactly $K \ll N$ of its entries are non-zero, i.e., $\|\boldsymbol{s}\|_0 \triangleq K$.*

Typically, not the signal $\boldsymbol{s}$ itself is sparse, but it admits a sparse representation in some basis $\boldsymbol{\Psi}$ such that

$$\boldsymbol{s} = \sum_{i=1}^{N} \boldsymbol{\psi}_i x_i = \boldsymbol{\Psi}\boldsymbol{x}, \|\boldsymbol{x}\|_0 = K \ll N, \tag{2.5}$$

where $\boldsymbol{\psi}_i$ are the columns of the matrix $\boldsymbol{\Psi}$. Common examples of such sparsifying transforms include Fourier and Wavelet [41]–[43]. The notion of sparsity is closely related to that of the signal support defined below.

Definition 2. *A support $\mathcal{S}(\boldsymbol{x})$ of some vector $\boldsymbol{x}$ is a set that contains positions of all its non-zero elements, i.e.,*

$$\mathcal{S}(\boldsymbol{x}) \triangleq \{i \in [1, N] : x_i \neq 0\} = \{s_1(\boldsymbol{x}), s_2(\boldsymbol{x}), \dots\}, \tag{2.6}$$

where $s_j(\boldsymbol{x})$ denotes the jth element of $\mathcal{S}(\boldsymbol{x})$.

Note that taking into account Definition 2, we can also write (2.5) as

$$\boldsymbol{s} = \sum_{p \in \mathcal{S}(\boldsymbol{x})} \boldsymbol{\psi}_p x_p, where \tag{2.7}$$

for any K-sparse vector $\boldsymbol{x}$ the cardinality of the support is $|\mathcal{S}(\boldsymbol{x})| = K$. Throughout the thesis, we will use the notation $\boldsymbol{x}_{\mathcal{S}(\boldsymbol{x})}$ to denote a vector of length $|\mathcal{S}(\boldsymbol{x})|$ that contains the elements of $\boldsymbol{x}$ indexed by $\mathcal{S}(\boldsymbol{x})$. Similarly, for some $M \times N$ matrix $\boldsymbol{B}$, the notation $\boldsymbol{B}_{\mathcal{S}(\boldsymbol{x})}$ will mean the $M \times |\mathcal{S}(\boldsymbol{x})|$ sub-matrix of $\boldsymbol{B}$ that contains the columns of $\boldsymbol{B}$ indexed by $\mathcal{S}(\boldsymbol{x})$.

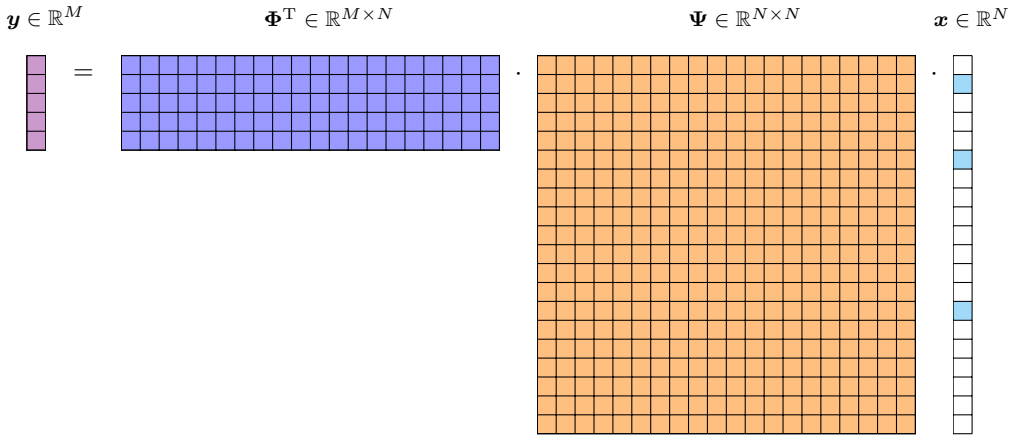


Figure 2.2: Finite-dimensional CS system with an M-length vector $\boldsymbol{y}$ containing measurements of a sparse (in a basis $\boldsymbol{\Psi}$) signal that is acquired via a measurement matrix $\boldsymbol{\Phi}$.

Now, we can substitute (2.5) into (2.4) to obtain

$$\boldsymbol{y} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{s} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}, \tag{2.8}$$

where $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$ is an $M \times N$ matrix that we will further refer to as the sensing matrix[2] that accounts for the measurement matrix $\boldsymbol{\Phi}$ and the sparsifying basis $\boldsymbol{\Psi}$. A schematic illustration of (2.8) is given in Figure 2.2. In case of (2.8), we are interested in exact recovery of $\boldsymbol{x}$ from $\boldsymbol{y}$, while we can always go back to $\boldsymbol{s}$ via (2.5) if needed.

In practice, it is rare for the signal of interest to be strictly sparse in the sense of Definition 1. Instead, the signal is more likely to be compressible, i.e., well-approximated by a limited number of coefficients [34], [44]. From the theory of non-linear approximation, it is known that the optimal (in terms of the ℓ_p-norm error) approximation of a signal $\boldsymbol{s} = \boldsymbol{\Psi}\boldsymbol{x}$ is obtained by simply setting the $(N - K)$ smallest coefficients x_k to zero [1]. As doing so yields again a K-sparse representation, for our purposes it suffices to consider the sparse model of (2.5).

Structured sparsity. Besides restricting the size of the support, the Definition 1 does not impose any further constraints on the positions of the non-zeros in $\boldsymbol{x}$. In many applications however, there are additional conditions that restrict possible supports to certain structures only. Such constraints can be expressed in a deterministic way or via some probability distribution over $\mathcal{S}(\boldsymbol{x})$. In both cases, explicit utilization of the underlying model can provide significant performance improvement over the standard sparse recovery algorithms [45]. An example of a common structure model is the so-called block sparsity where the non-zero coefficients appear in $\boldsymbol{x}$ in blocks of certain length or at regular intervals [36], [46], [47]. Another widespread type of structure is joint sparsity that is often exhibited when multiple observations of the same process are performed either simultaneously at different sensors

[2]Note that in CS and sparse recovery literature, both the sensing matrix $\boldsymbol{A}$ and the basis $\boldsymbol{\Psi}$ are commonly referred to as dictionaries and their columns as atoms.

[48] or over a short stationarity window in a single sensor [49]. The latter is known as the multiple measurement vector (MMV) model.

Parametric sparsity. Another important sparsity model that we consider here is the parametric model [34], [50]–[52]. As briefly mentioned at the beginning of Section 2.1, the signal $\boldsymbol{s}$ is often defined by a set of parameters $\{\theta_k\}_{k=1}^K \subset \Theta$ in some parameter space Θ such that

$$\boldsymbol{s} = \sum_{k=1}^{K} x_k \mathcal{F}(\theta_k), \tag{2.9}$$

where $\mathcal{F}$ represents a mapping from the parameter space $\Theta \subseteq \mathbb{R}$ to the signal space $\mathcal{S} \subset L_2$. The goal of solving (2.4) is to estimate $\{\theta_k\}_{k=1}^K$ from $\boldsymbol{y}$ rather than to recover the signal $\boldsymbol{s}$ or the coefficients $\boldsymbol{x}$. In this case, the matrix $\boldsymbol{\Psi}$ can be seen as a parametric dictionary that corresponds to some sampling $\{\theta_\ell^{(\mathrm{G})}\}_{\ell=1}^L \subseteq \Theta$ of the parameter space, i.e.,

$$\boldsymbol{\Psi} = [\mathcal{F}(\theta_1^{(\mathrm{G})}), \mathcal{F}(\theta_2^{(\mathrm{G})}), \ldots, \mathcal{F}(\theta_L^{(\mathrm{G})})]. \tag{2.10}$$

Assuming that the sampling grid[3] is chosen in such a way as to contain all θ_k, i.e., $\forall k \in [1, K]\ \exists \ell : \theta_\ell^{(\mathrm{G})} = \theta_k$, it holds that $\boldsymbol{s} = \boldsymbol{\Psi}\boldsymbol{x}$ and we can hence estimate $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^{\mathrm{T}}$ as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(\mathrm{G})}_{\mathcal{S}(\hat{\boldsymbol{x}})}, \tag{2.11}$$

where $\boldsymbol{\theta}^{(\mathrm{G})} = [\theta_1^{(\mathrm{G})}, \theta_2^{(\mathrm{G})}, \ldots, \theta_L^{(\mathrm{G})}]^{\mathrm{T}}$ is a vector that contains grid sampling points.

2.2.2 Sensing matrices

As we are familiar with the CS setup now, we can proceed with answering the first major CS question: how to choose the sensing matrix $\boldsymbol{A}$ such that $\boldsymbol{x}$ can be uniquely determined from $\boldsymbol{y}$? For clarity, we consider to this end the canonical (single measurement vector (SMV)) model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ that assumes[4] $\boldsymbol{\Psi} = \mathbf{I}_N$ where $\mathbf{I}_N$ denotes an $N \times N$ identity matrix; we will gradually include various other extensions as we advance towards more practical examples.

Null-space and matrix spark. To be able to recover all K-sparse vectors $\boldsymbol{x}$ from their linear measurements $\boldsymbol{A}\boldsymbol{x}$, we should choose $\boldsymbol{A}$ such as to ensure that any distinct K-sparse vectors $\boldsymbol{x} \neq \tilde{\boldsymbol{x}}$ have distinct measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{y}}$ [24], [34]. We can express this formally by introducing the null space of $\boldsymbol{A}$ as

$$\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{z} : \boldsymbol{A}\boldsymbol{z} = 0\}. \tag{2.12}$$

[3] This inherent dictionary gridding constitutes one of the biggest drawbacks of CS-based approaches, often challenging its practical applicability. We shortly review some of the problems it is associated with together with possible solutions in Section 2.2.5.

[4] As long as the basis $\boldsymbol{\Psi}$ is invertible and known when designing $\boldsymbol{\Phi}$, we can consider the canonical model without loss of generality since for $\boldsymbol{\Psi} \neq \mathbf{I}_N$ we can replace $\boldsymbol{\Phi}$ by $\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$ and achieve the same result.

Note that if $\mathcal{N}(\boldsymbol{A})$ contains any $2K$-sparse vectors then there is such a pair $\boldsymbol{x} \neq \tilde{\boldsymbol{x}}$ that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\tilde{\boldsymbol{x}}$ where both $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ are K-sparse. This is because the equality $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\tilde{\boldsymbol{x}}$ implies that $\boldsymbol{A}(\boldsymbol{x} - \tilde{\boldsymbol{x}}) = 0$, while one can always decompose a $2K$-sparse vector into two K-sparse ones. This observation leads to the condition on the null-space of $\boldsymbol{A}$ that guarantees a unique representation of any K-sparse vector $\boldsymbol{x}$: $\mathcal{N}(\boldsymbol{A})$ should not contain any vectors that are more than $(2K-1)$-sparse [34]. Another way of characterizing the above-mentioned condition is via the so-called matrix spark [10].

Definition 3. *The spark of a matrix $\boldsymbol{A}$ is the smallest number of its columns that are linearly dependent*[5]*:*

$$\text{spark}\,(\boldsymbol{A}) = \min_{d \neq 0} \|\boldsymbol{d}\|_0 \ \textit{s.t.}\ \boldsymbol{A}\boldsymbol{d} = 0. \tag{2.13}$$

We can now reformulate the null space condition on the identifiability of $\boldsymbol{x}$ using the matrix spark.

Theorem 1 (Corollary 1 of [10])**.** *For any vector $\boldsymbol{y} \in \mathbb{R}^M$ there is at most one K-sparse vector $\boldsymbol{x}$ such that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ if and only if* $\text{spark}\,(\boldsymbol{A}) > 2K$*.*

Theorem 1 immediately gives us a lower bound on the number of measurements required for unique recovery of K-sparse vectors. Note that for any $M \times N$ matrix $\boldsymbol{A}$, we have that $2 \leq \text{spark}\,(\boldsymbol{A}) \leq M + 1$ and therefore $M \geq 2K$.

The restricted isometry property. In [53], the authors introduce yet another property of $\boldsymbol{A}$ that plays an important role in establishing recovery guarantees in CS, namely the restricted isometry property (RIP).

Definition 4. *A matrix $\boldsymbol{A}$ is said to satisfy the RIP of order K if there is a constant $\delta \in (0, 1)$ such that for all K-sparse vectors $\boldsymbol{x}$ it holds that*

$$(1 - \delta)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1 + \delta)\|\boldsymbol{x}\|_2^2. \tag{2.14}$$

The constant δ in (2.14) is know as the RIP constant. The main idea behind (2.14) is that when $\boldsymbol{A}$ satisfies the RIP of order $2K$ and δ is reasonably small it means that $\boldsymbol{A}$ approximately preserves the Euclidean length of $2K$-sparse signals [18]. This in turn indicates that $\mathcal{N}(\boldsymbol{A})$ does not contain any $2K$-sparse vectors - exactly what we need for the identifiability of K-sparse signals from their measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Similarly, one can show that if $\boldsymbol{A}$ satisfies the RIP of order $2K$, then $\text{spark}\,(\boldsymbol{A}) > 2K$. Furthermore, [54] establishes that if $\delta \leq 0.5$ the number of measurements required for unique recovery is lower bounded by

$$M \geq CK \log\left(\frac{N}{K}\right), \tag{2.15}$$

where $C \approx 0.28$.

[5] By contrast, the rank of a matrix is the largest number of its columns that are linearly independent.

Coherence. The null-space condition, spark and the RIP all provide conditions on $\boldsymbol{A}$ that result in recovery performance guarantees. Ensuring that a given matrix $\boldsymbol{A}$ satisfies any of them is however a non-deterministic polynomial-time hard (NP-hard) problem as it typically requires performing a combinatorial search over all $M \times K$ sub-matrices of $\boldsymbol{A}$ [55]. An alternative measure whose computation is not as computationally demanding is the matrix coherence that determines the smallest (Euclidean) distance between two columns of a matrix [10], [56], [57].

Definition 5. *The coherence $\mu(\boldsymbol{A})$ of a matrix $\boldsymbol{A}$ is the largest inner product between its two columns $\boldsymbol{a}_i$, $\boldsymbol{a}_j$ that are normalized to their ℓ_2-norms, i.e.,*

$$\mu(\boldsymbol{A}) = \max_{i \neq j} \frac{|\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|}{\|\boldsymbol{a}_i\|_2 \|\boldsymbol{a}_j\|_2}. \tag{2.16}$$

In [10], it has been shown that (2.8) is guaranteed to have a unique K-sparse solution if

$$1 + \frac{1}{\mu(\boldsymbol{A})} > 2K. \tag{2.17}$$

The coherence is hence related to the spark as

$$\operatorname{spark}(\boldsymbol{A}) \geq 1 + \frac{1}{\mu(\boldsymbol{A})}, \tag{2.18}$$

whereas a (column) unit-norm $\boldsymbol{A}$ satisfies the RIP of order K with $\delta \leq (k-1)\mu(\boldsymbol{A})$ [10]. It is worth nothing that for a flat or a square matrix of size $M \times N$ with $N > 1$, the value of the coherence is lower[6] bounded by the so-called Welch bound [58], [59] such that

$$\mu(\boldsymbol{A}) \geq \sqrt{\frac{N-M}{M(N-1)}}. \tag{2.19}$$

Moreover, this bound is achieved only if inner product for all pairs of columns is of the same magnitude. Such matrices are called equiangular tight frames (ETF) [60].

Mutual coherence. For the composite model from Figure 2.2 one can also measure the largest distance between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ captured by the mutual coherence [18], [61].

Definition 6. *The mutual coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ between two matrices $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ of equal column dimension is given by*

$$\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \max_{i,j} \frac{|\langle \boldsymbol{\phi}_i, \boldsymbol{\psi}_j \rangle|}{\|\boldsymbol{\phi}_i\|_2 \|\boldsymbol{\psi}_j\|_2}. \tag{2.20}$$

From the definition, it is clear that $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ is large if $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ contain correlated columns, and it is small otherwise. Generally, in CS we aim at the maximally incoherent

[6] Clearly, the coherence is upper bounded by 1 since $|\langle \boldsymbol{a}_j, \boldsymbol{a}_i \rangle| \leq |\langle \boldsymbol{a}_i, \boldsymbol{a}_i \rangle|$ with the equality being true if and only if $\boldsymbol{a}_j = \boldsymbol{a}_i$ for some (i, j).

pairs $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ to allow for unique recovery [62]. For $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ with orthogonal columns, the lowest value of mutual coherence is determined by the column dimensionality as

$$\frac{1}{\sqrt{N}} \leq \mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \leq 1. \tag{2.21}$$

Suitable matrix constructions. As we have seen, there is a number of desired properties that a good sensing matrix should fulfill, spanning from a large spark and higher order RIP with sufficiently small constant to the low level of coherence. What remains open is how to construct $\boldsymbol{A}$ such that these are achieved.

By far, the largest class of matrices considered particularly suitable for CS are matrices whose elements are drawn from some probability distribution. More specifically, it has been shown that random matrices satisfy the RIP with high probability if their entries are drawn from sub-Gaussian distributions [14], [53], [63]–[65]. Common examples of sub-Gaussian ensembles include Gaussian, Bernoulli and Rademacher distributed random variables. More broadly, all bounded random variables fall into this category [64]. Random matrices also allow to achieve near optimal number of measurements in terms of (2.15) as well as low coherence that asymptotically reaches $\mu = \sqrt{\frac{2\log(N)}{M}}$ [41], [66]. Likewise, it is known that $M \times N$ matrices with i.i.d. elements have maximum spark of $M + 1$ with probability 1 [67]. Importantly, one can also show that in the setting of Figure 2.2, where $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$, a random construction of $\boldsymbol{\Phi}$ is likely to preserve these properties in $\boldsymbol{A}$ as well [63]. Yet another perspective on random matrix constructions as a choice for $\boldsymbol{\Phi}$ is their *universality*, i.e., they are likely to be mutually incoherent with any basis $\boldsymbol{\Psi}$. This in turn implies that by performing the measurement in a "(pseudo-)random" way one removes the need to know the sparsifying basis $\boldsymbol{\Psi}$ during the measurement step. In other words, with random measurements one should be able to reconstruct the signal for an arbitrary basis $\boldsymbol{\Psi}$, even if it changes meanwhile.

It is worth noting here that along with random matrices a number of deterministic constructions has been proposed over the years that fulfill one or more of the CS requirements. Thus, some $M \times M^2$ matrices achieve the lower bound on matrix coherence [59], [68], while Vandermonde matrices formed by distinct scalars are known to have maximal spark [69]. Possible constructions of full spark frames based on partial Fourier transform are also discussed in [70]. There are even ways to generate deterministic matrices that satisfy the RIP of a given order [71]–[77]. Despite their existence, these deterministic constructions often either have a very particular structure (as Vandermonde matrices for instance), which might lead to high coherence values, or result in a prohibitively large number of measurements. Additionally, they are significantly harder to analyze in the context of composite model $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$ as one has to explicitly consider the properties of the product between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$.

2.2.3 Noise models

The noise-free setting of (2.8) is seldom encountered in practice since the measurements are likely to be contaminated with noise. To account for this, we can write (2.8) as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{2.22}$$

where $\boldsymbol{n}$ is the additive noise vector that is typically modeled either as deterministic and bounded [11] or as white Gaussian [78], [79]. In CS literature, one comes across two types of noise often referred to as signal (or input) noise and measurement noise. To show the difference between the two we decompose $\boldsymbol{n}$ as

$$\boldsymbol{n} = \boldsymbol{A}\boldsymbol{n}_{\mathrm{s}} + \boldsymbol{n}_{\mathrm{m}}. \tag{2.23}$$

Here, $\boldsymbol{n}_{\mathrm{s}}$ and $\boldsymbol{n}_{\mathrm{m}}$ denote the signal and the measurement noise [80], respectively. From (2.23), the signal noise $\boldsymbol{n}_{\mathrm{s}}$ contaminates the signal prior to the measurement and undergoes the same measurement process as the signal. It can result from the physical hardware elements such as amplifiers, mixers and/or filters that are placed before the sampling. Contrarily, the measurement noise $\boldsymbol{n}_{\mathrm{m}}$ is added to the sampled signal $\boldsymbol{A}\boldsymbol{x}$. Hence, it accounts for the perturbations occurring during the measurement step such as quantization for instance.

Substituting (2.23) into (2.22), we obtain

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{n}_{\mathrm{s}} + \boldsymbol{n}_{\mathrm{m}} = \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{n}_{\mathrm{s}}) + \boldsymbol{n}_{\mathrm{m}}. \tag{2.24}$$

If not specified otherwise, we will assume throughout that $\boldsymbol{n}_{\mathrm{s}}$ and $\boldsymbol{n}_{\mathrm{m}}$ are independent random vectors with i.i.d zero-mean Gaussian distributed elements with variance σ_{s}^2 and σ_{m}^2, respectively.

2.2.4 Recovery algorithms

Now that we have learned under which conditions (2.8) has a unique K-sparse solution as well as several possible ways to ensure that these are fulfilled, the second set of questions we are concerned with is how to recover $\boldsymbol{x}$ from its measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

Basis Pursuit (ℓ_1 reconstruction). A straightforward way of approaching (2.8) is by searching for such solutions $\hat{\boldsymbol{x}}$ that have minimum ℓ_0-norm, i.e.,

$$\hat{\boldsymbol{x}} = \arg\left(\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0\right) \text{ s.t. } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{2.25}$$

Note that in the case of $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$, (2.25) simply transforms into

$$\hat{\boldsymbol{s}} = \boldsymbol{\Phi}^{\mathrm{T}}\hat{\boldsymbol{x}} : \ \hat{\boldsymbol{x}} = \arg\left(\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0\right) \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{x}. \tag{2.26}$$

The recovery strategy (2.25) is intuitive and it complies with the uniqueness conditions discussed above. However, its direct implementation is impractical since the ℓ_0-norm objective function is known to be non-convex and finding its (local) minimum is NP-hard [2], [81].

A major breakthrough in sparse signal processing was brought about by showing that, with the use of ℓ_1-norm instead of ℓ_0, (2.25) can be relaxed to a convex problem that can be then recast as a linear or quadratic, for which efficient solutions exist (i.e., with guaranteed polynomial runtime) [82]–[85]. In other words, we substitute (2.25) with the following minimization problem

$$\hat{\boldsymbol{x}} = \arg\left(\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1\right) \text{ s.t. } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{2.27}$$

known as the basis pursuit (BP). Note that this becomes possible because (2.27) is known to be the (tightest) convex relaxation of the original problem (2.25). In the presence of noise, (2.27) takes the form of the basis pursuit denoising (BPDN) [83]

$$\hat{\boldsymbol{x}} = \arg(\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1) \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \leq \varepsilon, \tag{2.28}$$

where ε is a constant that controls the data fidelity in terms of the mean squared error (MSE). The recovery guarantees for (2.27) and (2.28) have been intensively investigated from a variety of different perspectives with some of the fundamental results presented in [40], [62], [66], [79], [83], [86]–[89]. The bottom line here is that a sparse signal can be stably recovered using (2.28) in a number of noise settings from $M = \mathcal{O}(K \log N)$ measurements under the conditions similar to those discussed in Section 2.2.2, albeit somewhat tighter.

Greedy algorithms. As discussed above, optimization approaches[7] based on (2.28) are known to provide outstanding recovery performance under a wide range of circumstances. However, their complexity scales as $\mathcal{O}(N^3 \log(N))$ [96] which still might result in prohibitively long run-times for some practical applications [97]. This has lead to the development of lower complexity iterative algorithms that attempt at approximating (2.28). These are often categorized as greedy methods [95]. The name comes from the way the solution is obtained, by means of an iterative update of the signal support and the corresponding coefficients where at each step one or more components are identified [98].

To date, there is a large body of research dedicated to greedy methods with a wide variety of powerful algorithms available ranging from classical (general purpose) pursuits such as the matching pursuit (MP) [56] and the orthogonal matching pursuit (OMP) [57], [99]–[102] to more specialized algorithms specifically tailored to solving sparse recovery problems such as the compressive sampling matching pursuit (CoSaMP) [103] and the subspace pursuit (SP) [104] for instance. Thresholding algorithms such as iterative soft thresholding (IST) [105], [106] and iterative hard thresholding (IHT) [107]–[110] are also considered closely related. For more details on this and other methods we refer the reader to excellent overviews [95],

[7] Additionally to the BP type of formulation, these also include the Dantzig selector [90], [91], the least absolute shrinkage and selection operator (LASSO) [92]–[94] as well as several other [95].

Algorithm 1: Orthogonal Matching Pursuit [57]

Input: $\boldsymbol{A}$, $\boldsymbol{y}$

1 Initialize: $\hat{\boldsymbol{x}} = \boldsymbol{0}_{N\times 1}, \boldsymbol{r} = \boldsymbol{y}, \mathcal{S}(\boldsymbol{x}) = \emptyset, n_{\text{it}} = 0$

2 **while** *the stopping criterion is false* **do**

3 $\quad n_{\text{it}} = n_{\text{it}} + 1$;

4 $\quad \boldsymbol{b} = \boldsymbol{A}^{\text{T}}\boldsymbol{r}$;

5 $\quad \mathcal{S}(\boldsymbol{x}) = \mathcal{S}(\boldsymbol{x}) \cup \arg\max_{j=1\cdots N}(b_j)$;

6 $\quad \hat{\boldsymbol{x}}_{\mathcal{S}(\boldsymbol{x})} = \boldsymbol{A}^{\dagger}_{\mathcal{S}(\boldsymbol{x})}\boldsymbol{y}$;

7 $\quad \boldsymbol{r} = \boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}$;

Output: $\hat{\boldsymbol{x}}$

[98], [111] as well as the original publications [56], [99]–[101], [103]–[110]. Below we briefly describe one of the most popular sparse recovery algorithms, the OMP.

As sketched in Algorithm 1, OMP is a greedy algorithm that iteratively determines the support of $\boldsymbol{x}$ by finding the column in $\boldsymbol{A}$ that is most correlated with the measurements $\boldsymbol{y}$ (lines 4 and 5). Subsequently, it recovers the vector of active coefficients $\hat{\boldsymbol{x}}_{\mathcal{S}(\boldsymbol{x})}$ by solving a least-squares (LS) problem (line 6; $(\cdot)^{\dagger}$ here denotes the Moore-Penrose matrix pseudo-inverse) followed by subtracting the current estimate from $\boldsymbol{y}$ (line 7). This procedure is repeated n_{it} times until the stopping criterion is reached. The latter can be based on the value of the residual error or the number of iterations as, under certain conditions on $\boldsymbol{A}$, the OMP is known to be able to recover any K-sparse noise-free signal after exactly K iterations [57]. The recovery guarantees of OMP has been analyzed in [101] and more recently in [112] showing that it requires $M = \mathcal{O}(K^2 \log(N))$ measurements for stable reconstruction. Note that it has been also demonstrated that for a random $\boldsymbol{A}$ the OMP reaches the theoretical limit of $M = \mathcal{O}(K \log(N))$ with high probability [57]. The basic algorithm has seen a number of extensions, the most notable among which are the regularized [113] and the stagewise OMP [114].

Generally, greedy algorithms like OMP treat the sparse recovery task as a two-step problem; they first try to identify the support of $\boldsymbol{x}$ and then estimate the non-zero coefficients $\boldsymbol{x}_{\mathcal{S}(\boldsymbol{x})}$. Given $\mathcal{S}(\boldsymbol{x})$, the latter is straightforward and can be solved by standard LS as

$$\hat{\boldsymbol{x}}_{\Lambda} = \begin{cases} \boldsymbol{A}^{\dagger}_{\mathcal{S}(\boldsymbol{x})}\boldsymbol{y} & , \Lambda = \mathcal{S}(\boldsymbol{x}) \\ 0 & , \text{otherwise.} \end{cases} \tag{2.29}$$

The former is much more challenging which makes the support estimation performance one of the main (together with its complexity) characteristic of any practical recovery algorithm.

2.2.5 A note on gridding error

Adopting the signal model (2.8), we have discussed how a high dimensional signal that is sparse in some domain can be efficiently recovered from a few non-adaptive linear measurements. The model (2.8) implies that the signal is sparse in some finite-dimensional dictionary $\boldsymbol{\Psi}$, whereas in practice it is often characterized by a few parameters in a continuous parameter space. A common way to make the continuous model comply with the CS framework is by representing the continuous parameter space via its discrete version as in (2.10). Although it often results in satisfactory performance, such an approach is inherently troubled by the so-called gridding error problem. It manifests itself in the model mismatch that appears when the true signal parameters do not lie on the grid chosen for signal reconstruction [115], [116]. In order to mitigate the effects of the possible model mismatch one could think of increasing the density of the sampling grid making the gridding interval smaller and hence allowing for better signal approximation. This however results in the increase of the number of columns in $\boldsymbol{\Psi}$ and subsequently in $\boldsymbol{A}$ which i) increases the reconstruction complexity and ii) degrades the performance as the columns of $\boldsymbol{A}$ become more correlated [116]. This fundamental dilemma has triggered a substantial amount of research attempting to mitigate the effects of the gridding error. We briefly mention some of the promising ideas below, while noting that the gridding issue is outside the main scope of this thesis.

A first group of approaches to combat the gridding error is based on various grid adaptation methods that normally start with a fixed grid and proceed by iteratively refining it during the reconstruction. Examples of these include iterative grid refinement schemes [117]–[119] and space-alternating generalized expectation-maximization (SAGE)-like sparse estimation algorithms [120], [121]. Another group of approaches that could potentially be used to cope with the unknown grid positions includes dictionary learning methods where the sparsifying transform is learned together with the sparse signal [122]–[125]. Several works have also considered applying interpolation schemes to estimate the parameter value that falls in-between two neighboring dictionary atoms. Thus, in [126] and [127] schemes based on polynomial interpolation are considered, whereas in [128] these ideas are extended to include polar interpolation. Finally, an alternative way of approaching the sparse recovery problem that is based on the atomic norm (or total variation norm) minimization has been recently investigated in such works as [129]–[131]. It promises to enable a grid-less sparse recovery that eliminates the problem of discrete dictionary altogether [132].

2.3 Multiple measurement vector

The CS model that we have considered above is sometimes also referred to as the single measurement vector (SMV) as a counterpart to the multiple measurements version of (2.8) known under the name of the multiple measurement vector (MMV) [49], [133]–[136]. In the MMV setting, one acquires T measurements $\boldsymbol{y}_t \in \mathbb{R}^M$ of some sparse signals $\boldsymbol{x}_t$, $t \in [1, T]$.

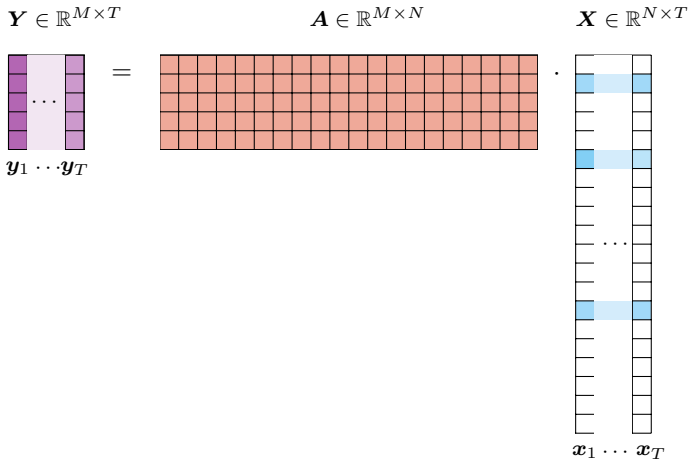


Figure 2.3: Finite-dimensional MMV CS system with an $M \times T$ matrix $\boldsymbol{Y}$ containing measurements of T K-sparse signals x_t that are acquired via the application of the sensing matrix $\boldsymbol{A}$.

The individual signals $\boldsymbol{x}_t$ can be collected locally at the same sensor [137] or from a number of different sensors in a distributed manner [48], [138]. In either case, all $\boldsymbol{x}_t$ are assumed to be jointly sparse such that that only K rows of $\boldsymbol{X}$ are non-identically zero. In other words, the support of $\boldsymbol{X}$, defined as

$$\mathcal{S}(\boldsymbol{X}) = \bigcup_{t=1}^{T} \mathcal{S}(\boldsymbol{x}_t), \tag{2.30}$$

satisfies $|\mathcal{S}(\boldsymbol{X})| = K \ll N$. The main premise of MMV is that, by recovering the jointly sparse $\boldsymbol{X}$ at once, one can improve the reconstruction performance compared to estimating each individual $\boldsymbol{x}_t$ independently. The MMV representation often appears in applications that involve observations of time-varying signals, such as the reception of radio frequency (RF) signals in wireless communications, radar, array processing, etc. Other examples include biomedical imagining where the scene is observed by a sensor that is moved around it and sensor networks with a number of spatially distributed sensors observing the same scene.

2.3.1 Data model

The MMV data model is formulated similarly to the SMV one, namely by introducing a sensing matrix $\boldsymbol{A}$ that performs dimensionality reduction such that

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}. \tag{2.31}$$

Here, $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T]$ is an $M \times T$ matrix each column of which contains a measurement $\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t$ of an (at most) K-sparse signal $\boldsymbol{x}_t$. The matrix $\boldsymbol{A}$ is an $M \times N$ sensing matrix with $N > M > K$ that has a similar meaning as in (2.8). Finally, a collection of vectors $\boldsymbol{x}_t$ forms a jointly K-sparse matrix $\boldsymbol{X}$ whose support satisfies (2.30), as illustrated in Figure 2.3.

2.3.2 Recovery guarantees and algorithms

The goal of the MMV problem associated with (2.32) is to recover the matrix $\boldsymbol{X}$ from the measurements $\boldsymbol{Y}$ for a known sensing matrix $\boldsymbol{A}$ [133], [136]. Following (2.25), we can formulate it as

$$\hat{\boldsymbol{X}} = \arg\left(\min_{\boldsymbol{X}} |\mathcal{S}(\boldsymbol{X})|\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}. \tag{2.32}$$

A sufficient and necessary condition for unique signal recovery in the MMV setting is [133]

$$|\mathcal{S}(\boldsymbol{X})| = K < \frac{\operatorname{spark}(\boldsymbol{A}) - 1 + \operatorname{rank}(\boldsymbol{Y})}{2}. \tag{2.33}$$

Note that, compared to the SMV case where $2K < \operatorname{spark}(\boldsymbol{A})$ (cf. Section 2.2.2), simultaneous recovery of multiple vectors enables reduction in the required number of measurements when $\operatorname{rank}(\boldsymbol{Y}) > 1$. One can also write (2.33) in terms of the rank of the sparse matrix $\boldsymbol{X}$ [136]:

$$2K < \operatorname{spark}(\boldsymbol{A}) - 1 + \operatorname{rank}(\boldsymbol{X}) \leq M + \operatorname{rank}(\boldsymbol{X}). \tag{2.34}$$

Doing so demonstrates that in the best-case scenario of $\operatorname{rank}(\boldsymbol{X}) = K$, only $K + 1$ measurements are sufficient for exact recovery of a K-sparse $\boldsymbol{X}$. Contrarily, in the worst case when the columns of $\boldsymbol{X}$ are colinear and $\operatorname{rank}(\boldsymbol{X}) = 1$, (2.34) becomes $2K < \operatorname{spark}(\boldsymbol{A})$ and MMV reduces to an SMV problem.

Optimization-based recovery. Analogously to the SMV, the joint sparse recovery problem (2.32) can be cast as an ℓ_0-norm optimization

$$\hat{\boldsymbol{X}} = \arg\left(\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{p,0}\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}, \tag{2.35}$$

where $\|\cdot\|_{p,q}$ denotes the $L_{p,q}$ matrix norm (as defined in Section 1.4). Relaxing the ℓ_0-norm in (2.35), one can attempt to solve a mixed norm optimization problem

$$\hat{\boldsymbol{X}} = \arg\left(\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{p,q}\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}, \tag{2.36}$$

where $p, q \geq 1$ [135]. Particularly, some popular choices for p and q include $1, 2$ and ∞. Note that the choice of $p = 1$ enforces the sparsity constraint, whereas $p = 2$ accounts for the total energy. From this perspective, for row-sparse matrices it is a good idea to choose $p = 1, q = 2$ while for column-sparse ones $p = 2, q = 1$ is a more natural choice.

Greedy methods. Many greedy algorithms originally introduced in the context of SMV sparse recovery can be immediately extended to accommodate the MMV model [134], [139]. This can be done for instance by replacing the vector correlation estimate $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}$ (line 6 of Algorithm (1)) by the row norms of the product between $\boldsymbol{A}$ and some residual matrix $\boldsymbol{R}$.

Algorithm 2: Simultaneous Orthogonal Matching Pursuit [134]

Input: $\boldsymbol{A}$, $\boldsymbol{Y}$

Initialize: $\hat{\boldsymbol{X}} = \boldsymbol{0}_{N\times T}, \boldsymbol{R} = \boldsymbol{Y}, \mathcal{S}(\boldsymbol{X}) = \emptyset, n_{\text{it}} = 0$

while *the stopping criterion is false* **do**

  $n_{\text{it}} = n_{\text{it}} + 1$;

  for $0 \leq i \leq N$ **do**

    $b_i = \|\boldsymbol{a}_i^{\text{T}}\boldsymbol{R}\|_q$;

  $\mathcal{S}(\boldsymbol{X}) = \mathcal{S}(\boldsymbol{X}) \cup \arg\max_{j=1\cdots N}(b_j)$;

  $\hat{\boldsymbol{X}}_{\mathcal{S}(\boldsymbol{X})} = \boldsymbol{A}^{\dagger}_{\mathcal{S}(\boldsymbol{X})}\boldsymbol{Y}$;

  $\boldsymbol{R} = \boldsymbol{Y} - \boldsymbol{A}\hat{\boldsymbol{X}}$;

Output: $\hat{\boldsymbol{X}}$

This is precisely what is done in the simultaneous orthogonal matching pursuit (SOMP), an MMV extension of the OMP that is summarized in Algorithm 2. An alternative approach, known as reduce MMV and boost (ReMBo)-OMP, is proposed in [137] where the MMV problem is first reduced to an equivalent SMV. Finally, it has been noticed that such direct extensions of SMV pursuits do not take advantage of the rank of $\boldsymbol{X}$ and hence have worse-case performance bounds close to those of the associated SMV problem [136]. This has triggered the development of a rank aware counterparts of SOMP such as rank aware orthogonal matching pursuit (RA-OMP) for example [136]. Interestingly, it turns out that when the rank of $\boldsymbol{X}$ is maximal, solving (2.32) does not require combinatorial complexity and the exact solution is guaranteed by simply applying such a well-known algorithm as multiple signal classification (MUSIC) [140] as demonstrated in [22], [136], [141]. More details on different MMV recovery strategies and performance analysis can be found in [98], [134], [135], [139], [142]–[145].

2.4 Sub-Nyquist sampling

With such a large amount of results available for discrete-dimensional CS models, the last important question that remains open is how to apply these to the actual problem at hand, the acquisition of sparse or compressible infinite-dimensional analog signals. There is likely to be no general answer to this question as different applications often impose significantly different physical constraints that define what is possible from the hardware point of view. However, for some signal classes, such as RF signals that are of main interest in the second part of this thesis, it is possible to outline a generic compressive receiver structure, while a particular realization will still always depend on a specific application [24]. The sub-Nyquist sampling framework that we introduce in the following is also known under the name of *Xampling* [39], [146]. Its two core components are the analog compression block whose aim is to reduce the analog bandwidth to be sampled and a non-linear detection block that

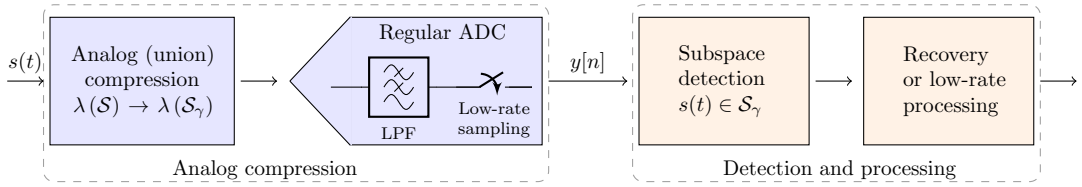


Figure 2.4: Xampling framework for compressive signal acquisition and processing of sparse analog signals. The figure is adapted from [39].

identifies active signal subspace after which classical signal processing algorithms can be used. This process is illustrated in Figure 2.4 where LPF stands for low-pass filter [39], [147].

Xampling. Xampling builds on the theory of shift-invariant (SI) subspaces, which is a special case of the union of subspaces model [35], [148], [149]. Generally, signals that belong to SI space can be expressed as a linear combination of some shifted *generator* functions $\tilde{\psi}_\ell(t)$ [148]

$$s(t) = \sum_{\ell=1}^{L} \sum_{n\in\mathbb{Z}} z_\ell[n]\tilde{\psi}_\ell(t - nT). \tag{2.37}$$

Here, $\tilde{\psi}_\ell(t)$ form a set of generators $\tilde{\Psi} = \{\tilde{\psi}_\ell(t)\}_{\ell=1}^{L}$, while $\boldsymbol{z}[n] = [z_1, \ldots, z_L[n]]^{\mathrm{T}}$ is a vector of coefficients z_ℓ for some time index n and a period T[35]. Alternatively, one can also consider (2.37) in the frequency domain by taking the Fourier transform of $s(t)$, i.e.,

$$S(f) = \sum_{\ell=1}^{L} \sum_{n\in\mathbb{Z}} \int_{-\infty}^{\infty} z_\ell[n]\tilde{\psi}_\ell(t - nT)e^{-\jmath 2\pi ft}\mathsf{d}t = \sum_{\ell=1}^{L} z_\ell(f)\psi_\ell(f), \tag{2.38}$$

where $\psi_\ell(f) \triangleq \int_{-\infty}^{\infty} \tilde{\psi}_\ell(t)e^{-\jmath 2\pi ft}\mathsf{d}t$ is the Fourier transform of $\tilde{\psi}_\ell(t)$ and $z_\ell(f) \triangleq \sum_n z_\ell[n]e^{-\jmath 2\pi fn}$ is the discrete time Fourier transform (DTFT) of $z_\ell[n]$ [35]. It is known that the signal of form (2.37) can be recovered from an output of a bank of L filters with almost [24, Chapter 3.1.1] arbitrary responses that are uniformly sampled at time instances nT [24], [35], [149].

When the signal is generated by only K out of L possible generator functions $\tilde{\psi}_\ell(t)$, we arrive at the sparse union of SI subspaces model (2.2). In other words, it means that in case of $K \ll L$ active generators at most K entries of $\boldsymbol{z}[n]$ (or $\boldsymbol{z}(f)$) are nonzero for each particular n (f in case of $\boldsymbol{z}(f)$). It has been demonstrated that it is possible to reduce the overall sampling rate in this case by producing the measurements as

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{z}[n], \tag{2.39}$$

where $\boldsymbol{A}$ is an $M \times L$ sensing matrix that allows recovering K-sparse signals [149]. In [24], [149], it is also shown that the sampling scheme that results in relation (2.39) can be obtained by passing $s(t)$ through M filters with appropriately chosen responses and uniformly sampling the outputs at nT, which provides the total sampling rate of M/T compared to L/T in the Nyquist case.

Expressions (2.37), (2.39) link the discrete CS model to sub-Nyquist sampling of continuous signals. Note that the system (2.39) is almost identical to the MMV discussed in Section 2.3, but for one important difference; it consists of an infinite number of measurements as $n \in \mathbb{Z}$. Accordingly, it is called the infinite measurement vector (IMV) [137]. Since $\|\boldsymbol{z}[n]\|_0 = K$ and $\boldsymbol{A}$ is chosen such as to comply with the CS requirements, one can attempt at recovering $\boldsymbol{z}[n]$ for each particular n independently[8]. However, this results in an infinite number of SMV systems to solve. Furthermore, in so doing one ignores the joint sparsity of $\boldsymbol{z}[n]$ along the time index n. The so-called continuous-to-finite (CTF) block [137] allows avoiding both of these issues by transforming (2.39) into a single MMV problem.

CTF block. To formulate the recovery of $\boldsymbol{z}[n]$ as an MMV problem, the CTF block first finds a basis for the linear span of $\boldsymbol{y}[n]$. One possible way to do so is to compute an $M \times M$ covariance matrix $\boldsymbol{Q}$ from the sequence of measurements $\boldsymbol{y}[n]$,

$$\boldsymbol{Q} = \sum_n \boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n], \tag{2.40}$$

and decompose it as $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}$, e.g., by means of the matrix eigen-decomposition for instance [24, Chapter 11.6.4]). This yields a frame $\boldsymbol{V}$ of finite dimensions whose column space is precisely the span of $\boldsymbol{y}[n]$. Given $\boldsymbol{V}$, we can then form a following MMV system

$$\boldsymbol{V} = \boldsymbol{A}\boldsymbol{U}, \tag{2.41}$$

where $\boldsymbol{U}$ is a jointly K-sparse matrix. The key point of obtaining (2.41) is that it can be shown that the support of its sparsest solution coincides with that of $\boldsymbol{z}[n]$ (see [149], [24, Chapter 11.6.4]) for more details). Therefore, by solving a regular finite-dimensional MMV problem (2.41), we can obtain the (row) support of the infinite-dimensional coefficient vector $\boldsymbol{z}[n]$, i.e., $\mathcal{S}(\boldsymbol{U}) = \mathcal{S}(\boldsymbol{z}[n])$. After the support is found, the non-zero coefficients $z_{\ell_k}[n] : \ell_k \in \mathcal{S}(\boldsymbol{z}[n])$ can be obtained by applying LS according to (2.29). If necessarily, one can further reconstruct the original continuous signal $s(t)$ by inserting $z_{\ell_k}[n]$ back into (2.37). This way, the CTF block bridges the final gap between the finite-dimensional CS theory considered so far and the sub-Nyquist sampling and recovery of sparse analog signals lying in a union of subspaces.

To conclude, we note that a possible practical application of (2.39) and (2.40) is examined in details in Part II of this thesis on the example of sub-Nyquist sampling and processing of analog multiband signals.

[8]Note that as along as one knows the set of generator functions $\{\tilde{\psi}_\ell(t)\}_{\ell=1}^L$, the recovery of $s(t)$ is tantamount to the recovery of the sequences $z_\ell[n]$ from $\boldsymbol{y}[n]$.

Chapter 3

Sensing matrix analysis and design

3.1 Motivation and related work

From the discussion in Chapter 2, we can see that the properties of the sensing matrix play a major role in establishing the recovery guarantees in CS, as well as analyzing what is practically achievable from the performance viewpoint in both noise-free and noisy settings. Furthermore, this applies more generally to any sparse estimation/recovery problem, irrespective of the model at hand. As mentioned, two particularly well-studied matrix fitness measures in CS are the matrix coherence [10], [56] and the RIP [53], [87]. Substantial amount of research is therefore dedicated to investigating different choices of the measurement kernel with respect to recovery bounds [62], [150]–[152]. As a result, random matrices with i.i.d. Gaussian or Bernoulli elements are commonly considered particularly suitable as they are likely to fulfill CS requirements [14], [53], [63]–[65]. In the presence of noise, the role of the sensing matrix becomes even more prominent as it can affect the sensitivity to noise [62] as well as the noise characteristics themselves [153], [154]. For instance, several papers have recently discussed the effects that the application of a sensing matrix has on the so-called input noise added to the signal prior the measurement [80], [155]. One of these is the coloring of the noise due to the non-orthogonality of the sampling functions. Another is the so-called noise folding effect which shows itself in the increase of the input noise power in the compressed measurements proportional to the compression ratio.

In this chapter, we investigate another important consequence of applying a compressive measurement kernel to the acquisition of sparse signals that has been largely overlooked so far, namely the variability of the output signal-to-noise ratio (SNR). It turns out that the effective signal power in compressed measurements depends on the entries of the sensing matrix corresponding to the support of the input signal. As a result, for a fixed input SNR, the effective SNR at the output of the receiver becomes dependent on the signal support; a situation that does not occur in traditional Nyquist rate sampling. This makes the SNR spread effect particularly important and calls for a better understanding of its extent and impact on the system performance. Therefore, the main goal of this chapter is to evaluate

the degree of such SNR variations in a greater detail and examine how it affects the recovery performance. In light of this, we begin the chapter with briefly discussing the effects of noise coloring and folding in Section 3.2. We then proceed to study the impact of the measurement kernel on the output SNR in Section 3.3 on the example of a canonical CS model with sensing matrices commonly used in CS, namely random Gaussian, Bernoulli and Rademacher matrices. To do so, we model the effective signal power as a random process whose characteristics are determined by the distribution from which the elements of the sensing matrix are drawn. Our analysis indicates that the coefficient of variation of the output SNR varies depending on the type of the sensing matrix used and it can be significant, especially for low number of measurements typical in CS. We further show that similar conclusions hold for some popular non-canonical CS model extensions as well.

This being said, intuitively one would wish that the effective system SNR is stable with respect to (non-power related) signal parameters and is close to that of the Nyquist rate system. However, what is of ultimate interest in CS system is the achievable recovery performance. Although the sparse recovery performance in CS has been extensively studied in a number of contexts including the worst-case [156], [157], the average case [139], [143] and the best-case [158] performance, the information-theoretic performance limits [159], [160] as well as probabilistic performance guarantees [152], [161], the specific issue of the support-dependent recovery performance, to the best of our knowledge, has not been discussed so far. We show however, that it is an inherent feature of the CS system that has a profound implications on the system design and performance. Therefore, in Section 3.4 we investigate the spread of the recovery performance with respect to the support of the input signal and the impact the output SNR has on it. We start by looking into the best case performance of an oracle-assisted estimator in terms of the mean squared error (MSE). This provides us with a Cramér-Rao bound (CRB)-type performance evaluation, which sets a benchmark for any practical recovery algorithm [158]. We then move to examining the performance of the support recovery. We do so by analyzing the probability of erroneous support estimation for correlation-based methods, both analytically and numerically. As a result, the performance in terms of both considered metrics proves to be highly non-uniform over the signal support. Furthermore, it clearly deteriorates for supports that correspond to lower output SNRs, which supports our hypothesis that the spread of the effective SNR negatively affects the overall system performance.

In order to account for the SNR spread and possibly other system parameters, we introduce a design framework that allows optimizing the measurement kernel according to specific system requirements. Note that while random matrix designs are known to fulfill a number of theoretical guarantees, drawing a measurement matrix from a random distribution has also several important disadvantages. One of these is the inability to exercise control over the properties of the resulting measurement kernel, which is particularly important when the measurement matrix has to be fixed to be implemented in hardware. The implementation of a random measurement kernel in a practical sensing system can be further restrained by

the hardware capabilities, application-specific requirements, operational limitations, etc. To illustrate some potential problems, consider a particular example of the Xampling sub-Nyquist sampling framework described in Section 2.4. The realization of a "random" measurement in this case would require constant change of the sensing functions (e.g., filters applied to $s(t)$). While the generation of pseudo-random sequences can be often done on the fly (think of sequence generation in spread spectrum communications for example [162]), in most practical systems one would still require to perform a calibration step to determine the exact entries of the measurement matrix, since they can differ significantly from the theoretically computed ones due to non-ideality of the hardware components. Failure to do so is likely to result in a model mismatch and, subsequently, a degradation of the recovery performance [163], [164]. Depending on the application, such a calibration step can be significantly time consuming, which makes a frequent change of the measurement kernel undesirable. The proposed design framework avoids these problems by providing a deterministic measurement design that also provides a possibility of including hardware-related requirements/constraints.

The material of this chapter is partially presented in [A1] and [A17]. An example of an application of the proposed measurement matrix design strategy to the compressive array design for direction of arrival (DoA) estimation can be found in [A3], [A6].

3.2 Impact on white noise: noise coloring and folding

Consider the canonical (real-valued) noisy CS model (2.24), i.e., $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{n}_{\mathrm{s}}) + \boldsymbol{n}_{\mathrm{m}}$. Under the assumption that both $\boldsymbol{n}_{\mathrm{s}}$ and $\boldsymbol{n}_{\mathrm{m}}$ are independent white Gaussian vectors whose elements have equal variances σ_{s}^2 and σ_{m}^2, respectively, the covariance matrix $\boldsymbol{\Sigma}$ of the total noise vector $\boldsymbol{n} = \boldsymbol{A}\boldsymbol{n}_{\mathrm{s}} + \boldsymbol{n}_{\mathrm{m}}$ can be written as

$$\boldsymbol{\Sigma} = \sigma_{\mathrm{s}}^2 \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} + \sigma_{\mathrm{m}}^2 \mathbf{I}_M, \tag{3.1}$$

Expression (3.1) shows a first consequence of the sensing matrix application in CS: the coloring of the signal noise $\boldsymbol{n}_{\mathrm{s}}$ by an arbitrary sensing matrix with $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \neq c\mathbf{I}_M$. Generally, the effect that the sensing matrix has on the covariance of the total noise vector $\boldsymbol{n}$ will depend on the noise ratio $\varsigma = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{m}}^2}$ and the difference $\delta_A = \min_c \|\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} - c\mathbf{I}_M\|_{\mathrm{F}}$, where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm.

In a special case when the rows of $\boldsymbol{A}$ are orthogonal with an equal squared norm[1] of $\frac{1}{\rho} = \frac{N}{M}$, the noise vector $\boldsymbol{n}$ stays white with covariance

$$\boldsymbol{\Sigma} = \frac{1}{\rho}\sigma_{\mathrm{s}}^2 \mathbf{I}_M + \sigma_{\mathrm{m}}^2 \mathbf{I}_M = \left(\frac{1}{\rho}\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{m}}^2\right)\mathbf{I}_M. \tag{3.2}$$

[1] This choice yields $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = N$, which corresponds to the Nyquist rate sensing when $\boldsymbol{A} = \mathbf{I}_N$.

In this case, the model (2.24) is equivalent to

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}_{\mathrm{b}}, \tag{3.3}$$

where $\boldsymbol{n}_{\mathrm{b}}$ is a white noise vector whose elements have variance $\sigma_0^2 = \frac{1}{\rho}\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{m}}^2$. From (3.2) we see that the variance of the signal noise after compression increases by the factor of $\frac{1}{\rho} = \frac{N}{M}$. This is due to the fact that the sensing matrix $\boldsymbol{A}$ combines the noise along the entire N-dimensional space whereas the signal itself resides in its (lower) K-dimensional sub-space. This results in an increase of the signal noise power in the compressed measurements that is inversely proportional to the compression ratio: the effect known as the *noise folding* [80], [155].

Interestingly, the models (2.24) and (3.3) can be shown to be roughly equivalent even in the case of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \neq c\mathbf{I}_M$ [155]. To illustrate this, let $\tilde{\boldsymbol{y}}$ be a whitened[2] version of $\boldsymbol{y}$ such that

$$\tilde{\boldsymbol{y}} = \sigma_0 \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} = \tilde{\boldsymbol{A}}\boldsymbol{x} + \tilde{\boldsymbol{n}}, \tag{3.4}$$

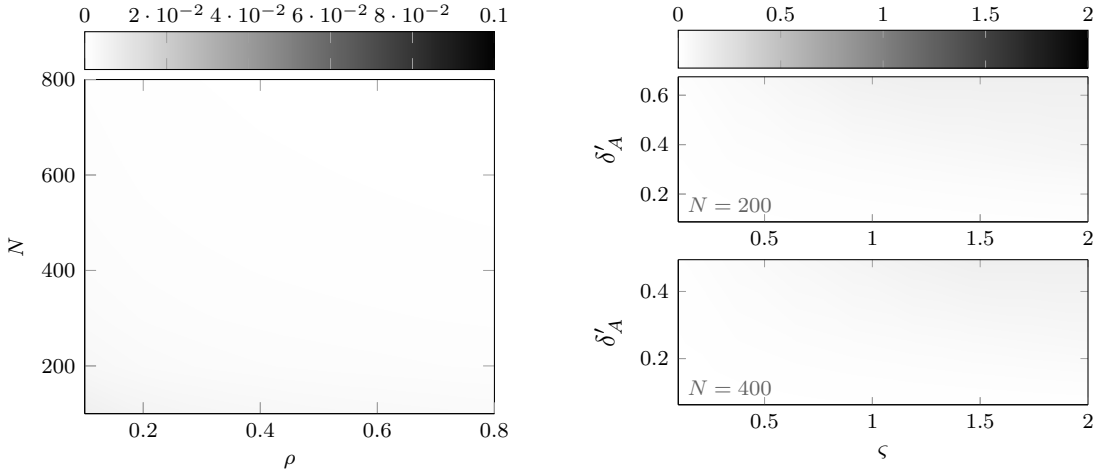
where $\tilde{\boldsymbol{A}} = \sigma_0 \boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}$ and $\tilde{\boldsymbol{n}} = \sigma_0 \boldsymbol{\Sigma}^{-1/2}\boldsymbol{n}$. Now, the noise vector $\tilde{\boldsymbol{n}}$ becomes white with covariance $\sigma_0^2\mathbf{I}_M$ as in (3.3). The remaining difference between (3.3) and (3.4) is the change of the sensing matrix from $\boldsymbol{A}$ to $\tilde{\boldsymbol{A}}$. In [155], it is shown that when δ_A is small, the coherence and the RIPs of $\tilde{\boldsymbol{A}}$ and $\boldsymbol{A}$ are very close. In other words, it implies that the recovery performance of (3.4) stays essentially unchanged compared to (3.3), provided that δ_A is small enough[3]. The same holds for $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$, where $\boldsymbol{\Phi}$ is drawn from sub-Gaussian distribution and $\boldsymbol{\Psi}$ is an orthogonal basis.

Example 3.1

As an example, Figure 3.1a shows the average difference δ_A/M as a function of the compression ratio $\rho = M/N$ and the signal dimensionality N for a random matrix $\boldsymbol{A}$ with i.i.d. elements drawn from a normal distribution $\mathcal{N}(0, \sqrt{1/M})$, further simply referred to as (random) Gaussian $\boldsymbol{A}$. It can be noted that δ_A decays rapidly with the increase of N and ρ, exhibiting significant values in the small area around the bottom left (low values) corner of the graph. Additionally, Figure 3.1b presents the deviation of the noise covariance matrix $\boldsymbol{\Sigma}$ calculated according to (3.1) from the covariance matrix of white noise $\boldsymbol{\Sigma}_{\mathrm{w}} = \left(\frac{1}{\rho}\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{m}}^2\right)\mathbf{I}_M$. The deviation is evaluated as δ_Σ/M where $\delta_\Sigma = \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\mathrm{w}}\|_{\mathrm{F}}$. Figure 3.1b demonstrates that under the prevailing measurement noise ($\varsigma < 1$) $\boldsymbol{\Sigma}$ can be well approximated by $\boldsymbol{\Sigma}_{\mathrm{w}}$. The same is true for larger values of N and ρ.

[2] In order to perform such a pre-whitening the noise covariance $\boldsymbol{\Sigma}$ has to be either known in advance or estimated beforehand, e.g., during a calibration step.

[3] Note that in the typical CS setting, when the elements of $\boldsymbol{A}$ are assumed to be drawn from sub-Gaussian distributions, δ_A is small with high probability [155].



(a) $\delta'_A = \frac{\delta_A}{M}$ **(b)** $\delta'_\Sigma = \frac{\delta_\Sigma}{M}$

Figure 3.1: Normalized difference $\delta'_A = \delta_A/M$ as a function of the compression ratio ρ and the dimension N (a) and normalized difference $\delta'_\Sigma = \delta_\Sigma/M$ as a function of δ'_A and $\varsigma = \frac{\sigma_s^2}{\sigma_m^2}$ (b).

3.3 Impact on signal power

3.3.1 SNR spread

Expressions (3.1)-(3.3) show that in the setting of (2.23) the application of the sensing matrix $\boldsymbol{A}$ can influence the characteristics of the (white) signal noise $\boldsymbol{n}_\mathrm{s}$. Consider now the influence of $\boldsymbol{A}$ on $\boldsymbol{x}$ and particularly on the measured signal power[4] $\|\boldsymbol{A}\boldsymbol{x}\|_2^2$, i.e.,

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \sum_{m=1}^{M} \left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k} x_k \right)^2 . \tag{3.5}$$

Expression (3.5) reveals another important effect arising from applying the sensing matrix $\boldsymbol{A}$ in (2.24), namely that the effective signal power in compressed measurements depends on the entries of the sensing matrix corresponding to the support of the input signal. Although this effect appears in both noise-free and noisy settings, its importance is better revealed in conjunction with the additive noise. Particularly, consider the output SNR (OSNR) [80]

$$\eta_\mathrm{O} \triangleq \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\mathbb{E}\left\{\|\boldsymbol{n}\|_2^2\right\}} = \frac{\|\boldsymbol{A}\boldsymbol{x}\|^2}{\mathbb{E}\{\|\boldsymbol{A}\boldsymbol{n}_\mathrm{s} + \boldsymbol{n}_\mathrm{m}\|_2^2\}} = \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{M\sigma_0^2} = \frac{\sum_{m=1}^{M} \left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k} x_k \right)^2}{M\sigma_0^2}, \tag{3.6}$$

where $\sigma_0^2 = \frac{1}{M}\mathrm{trace}\{\boldsymbol{A}\boldsymbol{A}^\mathrm{T}\}\sigma_\mathrm{s}^2 + \sigma_\mathrm{m}^2$. From (3.6), for fixed noise powers the OSNR η_O generally depends on the support $\mathcal{S}(\boldsymbol{x})$ via the corresponding values $a_{m,i}$ and the non-zeros x_i that we arrange into a sequence $\mathcal{X} = \{x_{i_1}, \ldots, x_{i_K}\}$ where $\forall k \in [1, K-1]\, i_k < i_{k+1} \in \mathcal{S}(\boldsymbol{x})$. The fact

[4] Note that we define the signal power here in deterministic terms: for a given sensing matrix $\boldsymbol{A}$ and a given input signal $\boldsymbol{x}$ with a certain support and certain values of its non-zero coefficients.

that the magnitudes of $\boldsymbol{x}$ have an impact on the SNR is not surprising as the SNR is meant to be a measure of the signal power with respect to the noise. What distinguishes (3.6) from the Nyquist rate settng, is that, other things being equal, the change of the signal support can lead to a change of the output SNR. As a result, the effective SNR at the output of the receiver might vary depending on the positions of the non-zeros in $\boldsymbol{x}$ leading to potentially non-uniform (over the support of the input signal) system performance. Importantly, this does not occur in the traditional Nyquist-rate sensing when $\boldsymbol{A} = \mathbf{I}_N$.

To illustrate the implications on the recovery performance, consider the recovered SNR (RSNR) [80] that accounts for the ratio of the signal power to the power of the reconstruction error caused by the presence of noise

$$\eta_{\mathrm{R}} \triangleq \frac{\|\boldsymbol{x}\|_2^2}{\mathbb{E}\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2\}}. \tag{3.7}$$

Naturally, the RSNR would largely depend on the particular algorithm used to solve (2.24). To circumvent this, we adopt an oracle-assisted approach to performance evaluation that assumes that the support of $\boldsymbol{x}$ is known prior to the recovery [80], [165], [166]. In doing so, we evaluate the best-case performance that corresponds to the case when a reconstruction algorithm has identified the support correctly, hence setting a benchmark for any practical recovery method. This said, when the true support $\mathcal{S}(\boldsymbol{x})$ is known, we obtain[5]

$$\eta_{\mathrm{R}} = \frac{\|\boldsymbol{x}\|_2^2}{\mathbb{E}\left\{\|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{y} - \boldsymbol{x}\|_2^2\right\}} = \frac{\|\boldsymbol{x}\|_2^2}{\mathbb{E}\left\{\|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{n}\|_2^2\right\}}. \tag{3.8}$$

Assuming for simplicity that the rows of $\boldsymbol{A}$ are orthogonal, we follow [80] and evaluate the performance of the oracle-assisted recovery via

$$\frac{\eta_{\mathrm{R}}}{\eta_{\mathrm{O}}} = \frac{\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{A}\boldsymbol{x}\|_2^2} \frac{\mathbb{E}\left\{\|\boldsymbol{n}\|_2^2\right\}}{\mathbb{E}\left\{\|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{n}\|_2^2\right\}} = \frac{\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{A}\boldsymbol{x}\|_2^2} \frac{M}{K} \|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\|_{\mathrm{F}}^{-1}. \tag{3.9}$$

Now, from the RIP we know that $\frac{1}{\sqrt{1+\delta}} \leq \frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{A}\boldsymbol{x}\|_2} \leq \frac{1}{\sqrt{1-\delta}}$ and from [54] that $\frac{1}{1+\delta} \leq \|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\|_{\mathrm{F}} \leq \frac{1}{1-\delta}$. Thus, the oracle-assisted RSNR is bounded by

$$\left(\frac{1-\delta}{1+\delta}\right) \frac{M}{K} \eta_{\mathrm{O}} \leq \eta_{\mathrm{R}} \leq \left(\frac{1+\delta}{1-\delta}\right) \frac{M}{K} \eta_{\mathrm{O}}. \tag{3.10}$$

Inequality (3.10) shows that for fixed δ, M and K the upper and lower bounds on the best-case RSNR scale linearly with the OSNR. This in turns indicates that the spread of the OSNR with respect to the signal support will result in a corresponding spread of the bounds on the RSNR. Therefore, one could expect a non-uniform (best-case) recovery performance over different signal supports. Being considered the most challenging part of sparse signal

[5]Note that $\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{y}$ is the unique solution to (2.8), provided that $\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}$ is full column-rank.

recovery, it is also reasonable to suspect then that the OSNR spread might have an even more dramatic impact on the support recovery performance.

Finally, in the presence of the signal noise $\boldsymbol{n}_{\mathrm{s}}$, one can also define an input SNR (ISNR)[6] which is the ratio of the total signal power prior the measurement to the input noise power:

$$\eta_{\mathrm{I}} \triangleq \frac{\|\boldsymbol{x}\|_2^2}{\mathbb{E}\{\|\boldsymbol{n}_{\mathrm{s}\mathcal{S}(\boldsymbol{x})}\|_2^2\}} = \frac{\|\boldsymbol{x}\|_2^2}{K\sigma_{\mathrm{s}}^2}. \tag{3.11}$$

Normalizing the output SNR η_{O} to the input SNR η_{I}, we obtain a measure of effective SNR change in comparison to the case of no compression:

$$\eta_{\mathrm{O/I}} = \frac{\eta_{\mathrm{O}}}{\eta_{\mathrm{I}}} = \frac{K\sigma_{\mathrm{s}}^2}{M\sigma_0^2}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} = \frac{K}{M}\frac{\varsigma}{\tau_{\boldsymbol{A}}\varsigma+1}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} = \frac{K}{M}\gamma\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2}, \tag{3.12}$$

where $\tau_{\boldsymbol{A}} = \frac{\mathrm{trace}\{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\}}{M}$ and $\gamma = \frac{\varsigma}{\tau_{\boldsymbol{A}}\varsigma+1}$ while $\varsigma = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{m}}^2}$. Since K, M and ς are constants while the value of $\tau_{\boldsymbol{A}}$ varies insignificantly[7] for random $\boldsymbol{A}$ that are appropriately normalized, the spread of η_{O} will also translate into a corresponding spread of $\eta_{\mathrm{O/I}}$. To evaluate the extent of such SNR variations, we investigate below the spread of the output SNR on the example of sensing matrices commonly used in CS.

3.3.2 Analytic analysis

We begin with examining the canonical CS model (2.24) where $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{n}$ and $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi} = \boldsymbol{\Phi}^{\mathrm{T}}$. Our goal is to characterize the spread of OSNR and its dependency on the elements of $\boldsymbol{A}$. From (3.6), η_{O} depends on the support of $\boldsymbol{x}$ via

$$\beta = \sum_{m=1}^{M}\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k\right)^2.$$

Note that in most applications the sensing matrix would be (at least temporarily) fixed after its elements are chosen (e.g., drawn according to some probability distribution). Hence, for a fixed $\boldsymbol{A}$ all $a_{m,n}$ are deterministic and for any given $\mathcal{X} = \{x_{i_1}, \ldots, x_{i_K}\}$ we can potentially determine a conditional frequency distribution $h_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$, as well as a sample mean and a sample variance of $(\eta_{\mathrm{O}}|\mathcal{X})$, by computing η_{O} for all possible supports $\mathcal{S}(\boldsymbol{x})$. The notation $h_{\boldsymbol{A}}(z)$ here indicates that the distribution of z is subject to change with the change of $\boldsymbol{A}$. To account for different $\mathcal{X}$, we can repeat this procedure for different combinations of signal magnitudes and average the results. This would result in a marginal frequency distribution $h_{\boldsymbol{A}}(\eta_{\mathrm{O}})$ over the support $\mathcal{S}(\boldsymbol{x})$. Although evaluating the spread of the output

[6] Alternatively to (3.11), one can define a *full-band* ISNR $\tilde{\eta}_{\mathrm{I}} \triangleq \frac{\|\boldsymbol{x}\|_2^2}{\mathbb{E}\{\|\boldsymbol{n}_{\mathrm{s}}\|_2^2\}}$. In case of white noise, $\tilde{\eta}_{\mathrm{I}} = \frac{K}{N}\frac{\|\boldsymbol{x}\|_2^2}{K\sigma_{\mathrm{s}}^2} = \frac{K}{N}\eta_{\mathrm{I}}$, which implies that the two differ by a factor of K/N only.

[7] From random matrix theory, the norms and the eigenvalues of an ensemble of random matrices tend to (asymptotically) concentrate around some mean, an effect known as concentration results [167].

SNR this way allows us to characterize a particular realization of $\boldsymbol{A}$, it might become computationally unfeasible as calculating $h_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ even for a single choice of magnitudes requires checking $C_N^K = \frac{N!}{K!(N-K)!}$ possible combinations. When the elements of $\boldsymbol{A}$ are drawn from some probability distribution and N is large enough, we can eliminate this difficulty by approximating the frequency distribution $h_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ by the (analytic) probability distribution $f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ derived by modelling the elements of $\boldsymbol{A}$ as i.i.d. random variables as explained in the following.

When $a_{m,n}$ are independently drawn from some probability distribution $f(\alpha)$, each row of $\boldsymbol{A}$ can be interpreted as containing an N-point sample from $f(\alpha)$. Re-writing (3.6) as

$$\eta_{\mathrm{O}} = \frac{1}{M\sigma_0^2}\beta = \frac{1}{M\sigma_0^2}\sum_{m=1}^{M} d_m^2, \tag{3.13}$$

where $\beta = \sum_{m=1}^{M} d_m^2$ and $d_m = \sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k$, we see that, for a given $\mathcal{X}$, d_m is then a linear combination of K realizations of some random variable $\alpha_m \sim f(\lambda)$. The support $\mathcal{S}(\boldsymbol{x})$ in this case defines which (K out of N) subset of realizations is taken. This enables the approximation of the frequency distribution $h_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ calculated for M particular sets of realizations $\{a_{m,n}\}_{n=1}^{N}$ by the probability distribution $f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ computed under the assumption that $\alpha_m \sim f(\alpha)$. Once $f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ is known, we can marginalize[8] $\mathcal{X}$ out to obtain

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}) = \int_{\mathcal{X}} f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})f(\mathcal{X})\mathsf{d}\mathcal{X} = \mathbb{E}_{\mathcal{X}}\{f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})\}, \tag{3.14}$$

where $\mathbb{E}_{\mathcal{X}}\{\cdot\}$ means the average over the ensemble of $\mathcal{X}$. Moreover, to evaluate the spread of η_{O} it is sufficient to calculate the mean $\mathbb{E}\{\eta_{\mathrm{O}}\} = \mathbb{E}_{\mathcal{X}}\{\mathbb{E}_{\boldsymbol{A}}\{\eta_{\mathrm{O}}|\mathcal{X}\}\}$ and the variance $\mathrm{var}\{\eta_{\mathrm{O}}\} = \mathbb{E}\{\eta_{\mathrm{O}}^2\} - \mathbb{E}^2\{\eta_{\mathrm{O}}\}$. This way, the knowledge of the distribution of $(\eta_{\mathrm{O}}|\mathcal{X})$ or of its mean and variance can be used to calculate the mean and variance of η_{O} for any prior on $\mathcal{X}$. Furthermore, analyzing $f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ allows us to separate the influence of $\boldsymbol{A}$ on the SNR spread from that of $\mathcal{X}$.

3.3.2.1 1-sparse input signals

We start with, we investigate the SNR spread in the special case of 1-sparse signals where

$$\eta_{\mathrm{O}} = \frac{\sum_{m=1}^{M} d_m^2}{M\sigma_0^2} = \frac{\sum_{m=1}^{M} a_{m,n}^2 x_n^2}{M\sigma_0^2} = \frac{x_n^2}{M\sigma_0^2}\|\boldsymbol{a}_n\|_2^2. \tag{3.15}$$

[8] We can do so for any deterministic sparsity pattern model including structured models such as the block sparsity for instance. Note that such sparsity models introduce additional (deterministic) constraints on the signal support that restrict the set of possible supports. Therefore, following the proposed approach, the analytic distribution $f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ will stay unchanged, whereas the approximation quality will deteriorate with the decrease of the support set size. On the other hand, imposing some probabilistic constraints on $\mathcal{S}(\boldsymbol{x})$ will require considering a joint distribution $f(\boldsymbol{A}, \mathcal{S}(\boldsymbol{x}))$ to derive $f_{\boldsymbol{A},\mathcal{S}(\boldsymbol{x})}(\eta_{\mathrm{O}}|\mathcal{X})$.

Here, $\boldsymbol{a}_n$ is the nth column of $\boldsymbol{A}$ whereas n is the index of the non-zero element in $\boldsymbol{x}$. Denoting by $\vartheta = \frac{\|\boldsymbol{x}\|_2^2}{M\sigma_0^2}$ the ratio of the (instantaneous) signal power to the total system noise power, η_{O} simply becomes

$$\eta_{\mathrm{O}} = \vartheta \|\boldsymbol{a}_n\|_2^2 = \vartheta\beta_1. \tag{3.16}$$

From (3.16), the spread of η_{O} over the index n is fully determined by that of $\beta_1 = \|\boldsymbol{a}_n\|_2^2$. Note that since η_{O} depends on $\mathcal{X}$ only via x_n^2, we formally have

$$\mathbb{E}\{\eta_{\mathrm{O}}\} = \mathbb{E}_{\mathcal{X}}\{\mathbb{E}_{\boldsymbol{A}}\{\vartheta\beta_1|\mathcal{X}\}\} = \mathbb{E}_{\boldsymbol{A}}\{\beta_1\mathbb{E}_{\mathcal{X}}\{\vartheta|\mathcal{X}\}\} = \vartheta_{\mathbb{E}}\mathbb{E}_{\boldsymbol{A}}\{\beta_1\}, \tag{3.17}$$

$$\mathrm{var}\{\eta_{\mathrm{O}}\} = \vartheta_{\mathbb{E}}^2\mathbb{E}_{\boldsymbol{A}}\{\beta_1^2\} - \vartheta_{\mathbb{E}}^2\mathbb{E}_{\boldsymbol{A}}^2\{\beta_1\} = \vartheta_{\mathbb{E}}^2\mathrm{var}_{\boldsymbol{A}}\{\beta_1\}, \tag{3.18}$$

where $\vartheta_{\mathbb{E}} = \frac{\mathbb{E}\{\|\boldsymbol{x}\|_2^2\}}{M\sigma_0^2} = \frac{P_{\mathrm{s}}}{M\sigma_0^2}$ with $P_{\mathrm{s}} = \mathbb{E}\left\{\|\boldsymbol{x}\|_2^2\right\}$ being the total signal power. In other words, when $K = 1$ the mean and the variance of η_{O} depend on $\mathcal{X}$ only via a constant $P_{\mathrm{s}} = \mathbb{E}\left\{x_n^2\right\}$. Finally, we notice that the distribution of $\eta_{\mathrm{O/I}}$ does not depend on $\mathcal{X}$ at all since

$$\eta_{\mathrm{O/I}} = \frac{K}{M}\frac{\varsigma}{\tau_{\boldsymbol{A}}\varsigma + 1}\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} = \frac{1}{M}\frac{\varsigma}{\tau_{\boldsymbol{A}}\varsigma + 1}\beta_1 = \frac{\gamma}{M}\beta_1. \tag{3.19}$$

Gaussian $\boldsymbol{A}$. Suppose the elements of $\boldsymbol{A}$ are drawn from a zero-mean normal Gaussian distribution such that $a_{m,i} \sim \mathcal{N}(0, 1/M)$. Then, $\beta_1 M$ is a random variable distributed according to the chi-squared distribution with M degrees of freedom, i.e., $\beta_1 M \sim \chi_M^2$.

Lemma 1 (See [168])**.** *Denote by $\Gamma(k, \theta)$ a Gamma distribution with a shape parameter k and a scale parameter θ. If $Y \sim \chi_L^2$ and c is a positive constant, then $cY \sim \Gamma\left(\frac{L}{2}, 2c\right)$.*

Proof. Cf. Appendix B.1.1. ■

From Lemma 1, we have that $\beta_1 \sim \Gamma\left(\frac{M}{2}, \frac{2}{M}\right)$ and hence

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X}) = \vartheta\Gamma\left(\frac{M}{2}, \frac{2}{M}\right) = \Gamma\left(\frac{M}{2}, \frac{2\vartheta}{M}\right). \tag{3.20}$$

The mean and variance of $(\eta_{\mathrm{O}}|\mathcal{X})$ are given by $k\theta = \vartheta$ and $k\theta^2 = \frac{2\vartheta^2}{M}$, respectively. From (3.17) and (3.18) we have that $\mathbb{E}\{\eta_{\mathrm{O}}\} = \mathbb{E}_{\mathcal{X}}\{\vartheta\} = \vartheta_{\mathbb{E}}$ and $\mathrm{var}\{\eta_{\mathrm{O}}\} = \vartheta_{\mathbb{E}}^2\mathrm{var}_{\boldsymbol{A}}\{\beta_1\} = \frac{2\vartheta_{\mathbb{E}}^2}{M}$.

Similarly, for $\eta_{\mathrm{O/I}}$ we immediately obtain

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}) = f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}) = \Gamma\left(\frac{M}{2}, \frac{2}{M}\frac{\gamma}{M}\right). \tag{3.21}$$

As mentioned at the end of Section 3.2, when the elements of $\boldsymbol{A}$ are drawn from sub-Gaussian distributions, δ_A is negligibly small. Therefore, $\mathrm{trace}\left\{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right\} \approx Mc = \frac{M}{\rho}$ and (3.21) can be approximated as

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}) \approx \Gamma\left(\frac{M}{2}, \frac{2}{MN}\frac{\varsigma}{\varsigma + \rho}\right). \tag{3.22}$$

Since the SNR is traditionally expressed on a logarithmic scale, the following proposition provides analogs of (3.20) and (3.21) for $\eta_{\mathrm{O}}^{\mathrm{L}} = 10\log_{10}(\eta_{\mathrm{O}})$ and $\eta_{\mathrm{O/I}}^{\mathrm{L}} = 10\log_{10}(\eta_{\mathrm{O/I}})$.

Proposition 1. *Let $EGa(\mu, \theta, k)$ denote the exponential-gamma distribution [169] with parameters μ, θ, and k whereas η_{O} and $\eta_{\mathrm{O/I}}$ are defined according to* (3.16) *and* (3.19)*, respectively. Then, for $\eta_{\mathrm{O}}^{\mathrm{L}} = 10\log_{10}(\eta_{\mathrm{O}})$ and $\eta_{\mathrm{O/I}}^{\mathrm{L}} = 10\log_{10}(\eta_{\mathrm{O/I}})$ we have that*

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}^{\mathrm{L}}) = EGa\left(k_{\mathrm{a}}\ln\left(\frac{2}{M}\right) + \vartheta_{\mathbb{E}}^{\mathrm{L}}, k_{\mathrm{a}}, \frac{M}{2}\right), \text{ and} \tag{3.23}$$

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}^{\mathrm{L}}) = EGa\left(k_{\mathrm{a}}\ln\left(\frac{2}{M}\right) + k_{\mathrm{a}}\ln\left(\frac{\gamma}{M}\right), k_{\mathrm{a}}, \frac{M}{2}\right), \tag{3.24}$$

where $\vartheta_{\mathbb{E}}^{\mathrm{L}} = 10\log_{10}(\vartheta_{\mathbb{E}})$ and $k_{\mathrm{a}} = \frac{10}{\ln(10)} \approx 4.34$.

Proof. Cf. Appendix B.1.2. ■

Note that by applying the same approximation as in (3.22) we can simplify (3.24) as

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}^{\mathrm{L}}) \approx \mathrm{EGa}\left(k_{\mathrm{a}}\ln\left(\frac{2}{M}\right) + k_{\mathrm{a}}\ln\left(\frac{\varsigma}{N(\varsigma+\rho)}\right), k_{\mathrm{a}}, \frac{M}{2}\right). \tag{3.25}$$

Bernoulli $\boldsymbol{A}$. Consider now a sensing matrix $\boldsymbol{A}$ whose elements are drawn from a Bernoulli distribution, i.e., $a_{m,k} \sim \mathcal{B}_e(p)$ with $p \in (0,1)$. In this case, $\beta_1 = \|\boldsymbol{a}_n\|_2^2$ follows a Binomial distribution of degree M with mean $\mathbb{E}\{\beta_1\} = Mp$ and variance $\mathrm{var}\{\beta_1\} = Mp(1-p)$. From (3.16), $(\eta_{\mathrm{O}}|\mathcal{X})$ is distributed according to a scaled Binomial distribution $\mathcal{B}_i^{\vartheta}(M,p)$ (see Appendix E.2) with mean $M\vartheta p$ and variance $M\vartheta^2 p(1-p)$, while the mean and variance of η_{O} are given by $M\vartheta_{\mathbb{E}}p$ and $M\vartheta_{\mathbb{E}}^2 p(1-p)$, respectively. Similarly, the ratio $\eta_{\mathrm{O/I}}$ follows a scaled Binomial distribution $\frac{\gamma}{M}\mathcal{B}_i^{\gamma/M}(M,p)$ with mean $p\gamma$ and variance $\frac{\gamma^2}{M}p(1-p)$.

For the logarithmic scale, the probability mass functions (PMFs) of $(\eta_{\mathrm{O}}^{\mathrm{L}}|\mathcal{X})$ and $\eta_{\mathrm{O/I}}^{\mathrm{L}}$ do not strictly exist as the set of values of η_{O} includes 0 for which the logarithm is not defined. Nevertheless, by excluding the zero point we can still use the notion of PMF, albeit with a slight abuse of notation[9]. Then, we have that for $m = 1, 2, \cdots, M$

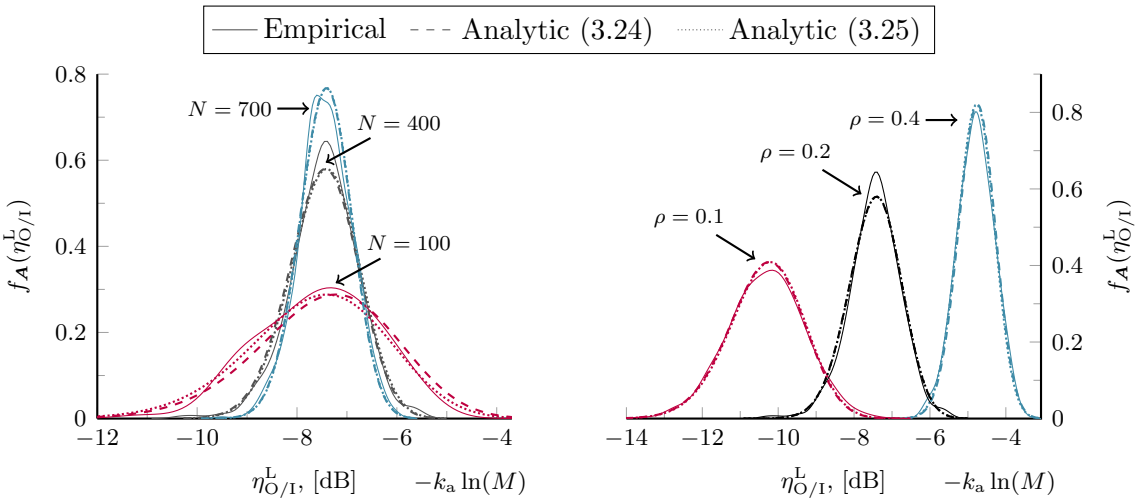
$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}^{\mathrm{L}}|\mathcal{X}) \approx \Pr\left[\eta_{\mathrm{O}}^{\mathrm{L}} = \vartheta^{\mathrm{L}} + k_{\mathrm{a}}\ln(m)\right] = \mathrm{C}_m^M p^m (1-p)^{M-m}, \text{ and} \tag{3.26}$$

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}^{\mathrm{L}}) \approx \Pr\left[\eta_{\mathrm{O/I}}^{\mathrm{L}} = k_{\mathrm{a}}\ln\left(\frac{\gamma}{M}\right) + k_{\mathrm{a}}\ln(m)\right] = \mathrm{C}_m^M p^m (1-p)^{M-m}. \tag{3.27}$$

Example 3.2

Figure 3.2 demonstrates the empirical and the analytic (according to expressions (3.24) and (3.25)) probability density functions (PDFs) of $\eta_{\mathrm{O/I}}^{\mathrm{L}}$ for a single realization of a

[9] The resulting function is not a probability measure anymore because it does not sum to 1. However, the difference can be considered practically negligible as the missing term $(1-p)^M$ decays exponentially with M.



(a) $\rho = 0.2$, $\varsigma = 1$ **(b)** $N = 400$, $\varsigma = 1$

Figure 3.2: Empirical and analytic PDFs of $\eta^{\mathrm{L}}_{\mathrm{O/I}}$ for a 1-sparse signal and a single realization of a Gaussian $\boldsymbol{A}$: a fixed ρ and several values of N (a), and a fixed N and several values of ρ (b).

random Gaussian $\boldsymbol{A}$ with different values of N (Figure 3.2a) and $\rho = M/N$ (Figure 3.2b). In both figures, the noise ratio ς is fixed and equal to 1 while the x-axes are scaled as $x + k_{\mathrm{a}} \ln(M)$ in order to highlight the noise folding effect whose contribution to the relative SNR deterioration is roughly proportional to $k_{\mathrm{a}} \ln(\rho)$.

Evidently, both analytic expressions provide a good fit to the empirical distribution of the relative OSNR over the signal position. The deviation grows with the decrease of M and N as the size of $\boldsymbol{A}$ becomes smaller. Furthermore, we see that the SNR variations for considered parameter settings can be as much as $2 - 5$ dB, with larger values for smaller M.

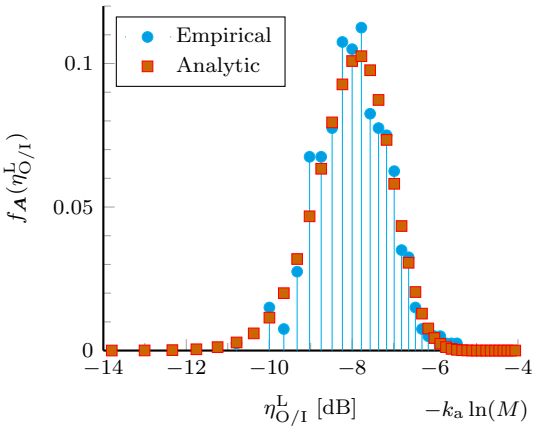


Figure 3.3: Empirical and analytic PMF of a 1-sparse signal and a signal realization of a scaled Bernoulli $\boldsymbol{A}$ with $N = 400$, $\rho = 0.2$ and $\varsigma = 1$.

Figure 3.3 in turn shows the empirical and analytic PMFs of $\eta_{\mathrm{O/I}}$ for a Bernoulli $\boldsymbol{A}$ with the probability of success $p = 0.5$ where $N = 400$, $\rho = 0.2$ and $\varsigma = 1$. For a fair comparison with the Gaussian sensing matrix of the same dimensions, the elements of $\boldsymbol{A}$ are additionally scaled by $\sqrt{\frac{1}{pM}}$ in order for the row norm to reach $\frac{1}{\rho}$. Note that such a scaling does not impact the form of the distribution, but only scales its range of values. From Figure 3.3, we can see that the range and the spread of the OSNR for a Bernoulli $\boldsymbol{A}$ is comparable to that of its Gaussian counterpart from Figure 3.2.

Rademacher $\boldsymbol{A}$. The last sensing matrix type we consider here is the Rademacher $\boldsymbol{A}$, whose elements take value ± 1 with probability 0.5. Obviously, for such a matrix the ℓ_2 norm of its columns is constant, resulting in a deterministic output SNR equal to $\frac{P_\mathrm{s}}{\sigma_0^2}$. The ratio $\eta_{\mathrm{O/I}} = \gamma$ is then also constant and independent of the signal position. This means that in the case of 1-sparse signals, the Rademacher sensing matrix provides a constant OSNR whose value is defined solely by ISNR, ς and $\tau_{\boldsymbol{A}}$. In the following we will see however, that this property is not preserved for the more general case of $K > 1$.

3.3.2.2 Arbitrary-sparse input signals

Untill now, we have considered a restricted class of input signals that have only a single non-zero element. Thanks to the resulting linear relation between the OSNR and the ℓ_2-norm of the sensing matrix columns, we have been able to analytically describe the spread of the OSNR in this case. Such an analysis is helpful in providing an initial insight into the SNR variability for a particular choice of the sensing matrix. However, it might be insufficient for providing a conclusive judgment. Therefore, below we examine a generic case of $K > 1$.

Gaussian $\boldsymbol{A}$. For a random Gaussian $\boldsymbol{A}$ and $K > 1$, $d_m = \sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k} x_k$ is a zero-mean normal variable with variance $\sigma_m^2 = \sum_{k=1}^{K} {x_k}^2 / M = \|\boldsymbol{x}\|_2^2 / M$. Recalling that $\beta = \sum_{m=1}^{M} d_m^2$, we invoke Lemma 1 to obtain

$$f_{\boldsymbol{A}}(\beta|\mathcal{X}) = \Gamma\left(\frac{M}{2}, \frac{2\sigma_m^2}{M}\right), \text{ and} \tag{3.28}$$

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X}) = \Gamma\left(\frac{M}{2}, \frac{2\sigma_m^2}{M^2\sigma_0^2}\right) = \Gamma\left(\frac{M}{2}, \frac{2\vartheta}{M}\right). \tag{3.29}$$

Lemma 2. *Let $\boldsymbol{x}$ be an arbitrary K-sparse vector of length N and $y \sim \Gamma\left(\frac{M}{2}, \frac{2\|\boldsymbol{x}\|_2^2}{M}\right)$. Then for any $M > 0$ we have that $\mathbb{E}\{y\} = P_\mathrm{s}$ and $\mathrm{var}\{y\} = \frac{2}{M} P_\mathrm{s}^2$.*

Proof. Cf. Appendix B.1.3. ■

From Lemma 2, $\mathbb{E}\{\eta_{\mathrm{O}}\} = \vartheta_{\mathrm{E}}$ and $\mathrm{var}\{\eta_{\mathrm{O}}\} = \frac{2}{M}\vartheta_{\mathrm{E}}^2$, while $\mathbb{E}\left\{\eta_{\mathrm{O/I}}\right\} = \frac{1}{M}\gamma$ and $\mathrm{var}\left\{\eta_{\mathrm{O/I}}\right\} = \frac{2\gamma^2}{M^3}$. These results coincide with the ones obtained for a Gaussian $\boldsymbol{A}$ with $K = 1$ showing that the OSNR spread in case of Gaussian $\boldsymbol{A}$ is independent of K.

Bernoulli $\boldsymbol{A}$. When $a_{m,n} \sim \mathcal{B}_e(p)$, each $a_{m,k}x_k$ is a random variable distributed as scaled Bernoulli where the scaling depends on the signal values. The distribution of the K-term sum of such random variables can be described directly via probabilities as

$$\Pr\left[d_m = \sum_{k=1}^{K} b_{n,k} x_k | \mathcal{X}\right] = p^n (1-p)^{K-n}, \tag{3.30}$$

where $b_{n,k} \in \{0,1\}$ is a kth element of a K-length binary vector $\boldsymbol{b}_n$ that contains exactly $n \in [0,K]$ ones[10]. From (3.30) it becomes obvious that, for a given $\mathcal{X}$, $\beta = \sum_{m=1}^{M} d_m^2$ does not converge to the Binomial distribution anymore. Generally, there is no simple way to describe the distribution of β. However, in the special case when all non-zero elements of $\boldsymbol{x}$ are equal (i.e., $x_k = x \; \forall k \in [1,K]$), $d_m = x \sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k}$ again becomes scaled Binomial distributed. In this case, we can calculate the mean and variance of d_m^2 as (see App. B.1.4 for details)

$$\mathbb{E}\{d_m^2\} = \big((K-1)p+1\big)pP_s \tag{3.31}$$

$$\operatorname{var}\{d_m^2\} = \frac{1+2p\big(K-1\big)\big((2K-3)p+3\big)}{K}(1-p)pP_s^2. \tag{3.32}$$

From (3.31) and (3.32), we then obtain

$$\mathbb{E}\left\{\eta_{\mathrm{O/I}}\right\} = \frac{\gamma}{M} M \mathbb{E}\left\{d_m^2\right\} = \gamma p(1+(K-1)p)K, \tag{3.33}$$

$$\operatorname{var}\left\{\eta_{\mathrm{O/I}}\right\} = \left(\frac{\gamma}{M}\right)^2 M \operatorname{var}\left\{d_m^2\right\} = \frac{\gamma^2}{M}\Big(1+2p\big(K-1\big)\big((2K-3)p+3\big)\Big)K(1-p)p. \tag{3.34}$$

Expressions (3.33)-(3.34) show that in the case of Bernoulli $\boldsymbol{A}$ and an input signal with non-zero entries of equal magnitudes both the mean and the variance of the relative output SNR grow with the increase of the number of non-zeros in $\boldsymbol{x}$.

Rademacher $\boldsymbol{A}$. A similar situation occurs with the Rademacher distributed $\boldsymbol{A}$ where each $a_{m,k}x_k$ can take the value of x_k or $-x_k$ with equal probability. Indeed, since $d_m = \sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k}x_k$, the distribution of d_m^2 depends on the particular values of x_k and hence is tedious to formulate. However, in the special case when all x_k are equal, we have that $d_m^2 = x^2\left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k}\right)^2$ which significantly simplifies further analysis. More specifically, since each $a_{m,k} = \pm 1$, the squared sum of K of them is a random variable that takes integer values in a set $\{K-2n\}_{n=0}^{\lfloor 0.5K \rfloor}$, where $\lfloor \cdot \rfloor$ denotes the operation of rounding to the nearest lower integer. The parameters of this random variable are described by the following lemma.

Lemma 3. *Let a_k be i.i.d. random variables that take values ± 1 with equal probability and $\omega = \left(\sum_{k=1}^{K} a_k\right)^2$. Then, the PMF of ω is given by*

$$\begin{cases} Pr\left[\omega = (K-2n)^2\right] = 2C_n^K 2^{-K}, & n = 0,1,\cdots,\lfloor 0.5K \rfloor - 1 \\ Pr\left[\omega = (K-2\lfloor 0.5K \rfloor)^2\right] = \left(K_{\mathrm{mod}_2}+1\right) C_n^K 2^{-K}, & n = \lfloor 0.5K \rfloor, \end{cases} \tag{3.35}$$

where mod_n *denotes the modulo n operation. Furthermore, the expected value of ω is* $\mathbb{E}\{\omega\} = K$ *and the variance is* $\operatorname{var}\{\omega\} = 2K(K-1)$.

Proof. Cf. Appendix B.1.5. ■

[10] Note that for each n there exists C_n^K of different vectors $\boldsymbol{b}_n$ that differ by the locations of ones and zeros.

Applying Lemma 3 to $d_m^2 = x^2\omega$ and taking into account that $\mathbb{E}\left\{d_m^2\right\} = \mathbb{E}_{\mathcal{X}}\{x^2\mathbb{E}\{\omega\}\} = \frac{P_{\mathrm{s}}}{K}\mathbb{E}\{\omega\}$ and $\mathrm{var}\left\{d_m^2\right\} = \mathbb{E}_{\mathcal{X}}\{x^4\mathbb{E}_{\boldsymbol{A}}\{\omega^2\}\} - \frac{P_{\mathrm{s}}^2}{K^2}\mathbb{E}\{\omega\}^2 = \frac{P_{\mathrm{s}}^2}{K^2}\mathrm{var}\{\omega\}$, we immediately obtain

$$\mathbb{E}\left\{\eta_{\mathrm{O/I}}\right\} = \frac{K\sigma_{\mathrm{s}}^2}{P_s}\frac{MP_s}{M\sigma_0^2} = K\gamma, \tag{3.36}$$

$$\mathrm{var}\left\{\eta_{\mathrm{O/I}}\right\} = \frac{P_s^2}{M\sigma_0^4}\frac{K^2\sigma_{\mathrm{s}}^4}{P_s^2}\frac{2(K-1)}{K} = \frac{\gamma^2}{M}2(K-1)K. \tag{3.37}$$

Note that for $K > 1$, the OSNR spread depends on K.

3.3.2.3 Coefficient of variation

In order to evaluate the OSNR spread for different values of K we employ the coefficient of variation also known as the relative standard deviation.

Definition 7. *For some random variable z, the coefficient of variation is defined as the square root of the ratio of its variance to the square of its expected value, i.e.,*

$$c_{\mathrm{v}}(z) = \sqrt{\frac{\mathrm{var}\{z\}}{\mathbb{E}^2\{z\}}}, \tag{3.38}$$

Given (3.38), the coefficient of variation for three considered above cases can be computed as

1. Gaussian $\boldsymbol{A}$: $c_{\mathrm{v}}(\eta_{\mathrm{O}}) = \sqrt{\frac{2}{M}}$;
2. Bernoulli $\boldsymbol{A}$ with $x_k = x\ \forall k = 1,\ldots,K$: $c_{\mathrm{v}}(\eta_{\mathrm{O}}) = \sqrt{\frac{1}{M}\frac{1+2p\left(K-1\right)\left((2K-3)p+3\right)}{K\left((K-1)p+1\right)^2 p}(1-p)}$, where for $p = 0.5$ it simplifies to $c_{\mathrm{v}} = \sqrt{\frac{1}{M}\frac{2K-1}{2K(K+1)}}$;
3. Rademacher $\boldsymbol{A}$ with $x_k = x\ \forall k = 1,\ldots,K$: $c_{\mathrm{v}}(\eta_{\mathrm{O}}) = \sqrt{\frac{1}{M}\frac{2(K-1)}{K}}$.

Since $\eta_{\mathrm{O/I}}$ differ from η_{O} only by a multiplicative constant, we have that $c_{\mathrm{v}}(\eta_{\mathrm{O}}) = c_{\mathrm{v}}(\eta_{\mathrm{O/I}})$.

Example 3.3

In this example, we show in Figure 3.4 the analytic coefficient of variation for three considered types of sensing matrices as a function of the sparsity level K. To be independent of the system parameters, it is normalized to $\sqrt{1/M}$. We can see that both Gaussian (denoted by $\mathcal{N}$) and Rademacher (denoted by $\mathcal{R}$) sensing matrices provide a higher coefficient of variation than a Bernoulli (denoted by $\mathcal{B}_e(p)$) one, with the notable exception of the Rademacher $\boldsymbol{A}$ for the case of a 1-sparse signal. Furthermore, while the coefficient of variation of the Rademacher sensing matrix grows with the increase of K, the Bernoulli coefficient of variation declines.

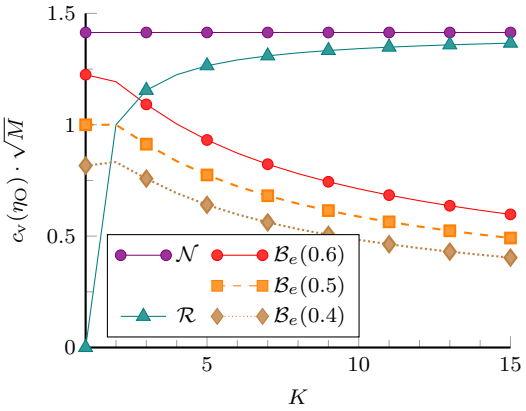
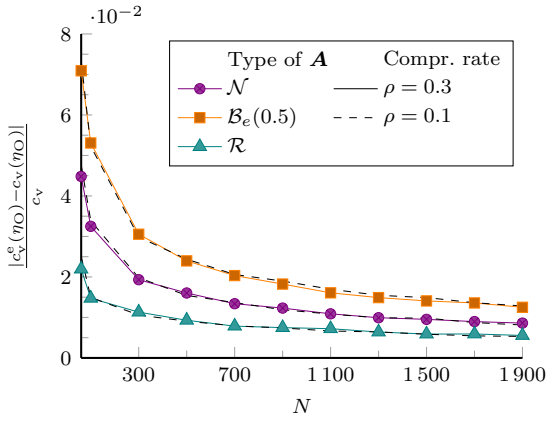


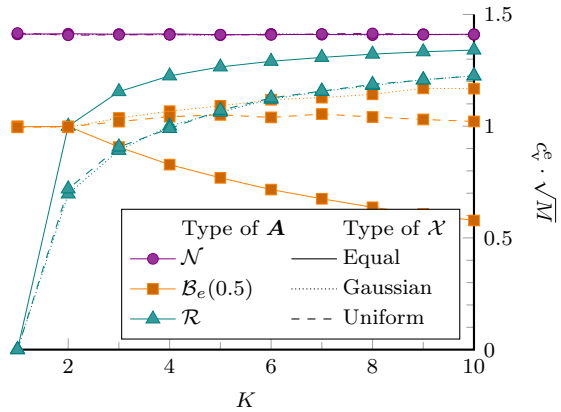
Figure 3.4: Analytic coefficient of variation c_{v} as a function of the sparsity level K.

To validate the derived analytic expressions, we calculate the average (among realizations of the sensing matrix) root mean squared error (RMSE) between the coefficients of variation obtained numerically ($c_{\mathrm{v}}^{\mathrm{e}}$) and analytically ($c_{\mathrm{v}}$). Figure 3.5a displays the normalized RMSE as a function of the signal dimensionality N for $K = 2$ and two values of the compression ratio ρ. In both cases, the RMSE does not exceed 7% of the predicted value with the best correspondence exhibited by the Rademacher $\boldsymbol{A}$. As expected, the error decays as N grows due to the increasing accuracy of the analytic approximation.

To this end, for Bernoulli and Rademacher sensing matrices we have considered the case when the input signal entries are of equal magnitude. This is because the rigorous analytic analysis of the generic case for these two types of sensing matrices is severely hampered by the dependency of the OSNR distribution on the values of the non-zero signal entries. To investigate how the choice of $\mathcal{X}$ influences the coefficient of variation, Figure 3.5b shows average (among 10^3 realizations of $\boldsymbol{A}$ and $\mathcal{X}$) empirical coefficient of variation $c_{\mathrm{v}}^{\mathrm{e}}$ as a function of K computed for three considered types of $\boldsymbol{A}$ and different priors on $\mathcal{X}$. Namely, i) all x_i



(a) RMSE $\frac{|c_{\mathrm{v}}^{\mathrm{e}}(\eta_{\mathrm{O}}) - c_{\mathrm{v}}(\eta_{\mathrm{O}})|}{c_{\mathrm{v}}}$



(b) Coefficient of variation $c_{\mathrm{v}}^{\mathrm{e}}$

Figure 3.5: Normalized RMSE between the empirical and the analytic coefficient of variation as a function of N with $K = 2$ (a). Normalized average $c_{\mathrm{v}}^{\mathrm{e}}$ for equal (solid lines), Gaussian (dotted lines), and Uniform (dashed lines) signal magnitudes with $N = 300$ and $\rho = 0.1$ (b).

have equal magnitudes; ii) x_i are i.i.d zero-mean normal random variables with variance $1/K$; iii) x_i are i.i.d zero-mean random variables uniformly distributed on the interval[11] $[-\sqrt{3/K}, \sqrt{3/K}]$. As expected, the coefficient of variation for a Gaussian $\boldsymbol{A}$ depends on neither the value of K nor the type of $\mathcal{X}$ and it is equal to $c_{\mathrm{v}} \cdot \sqrt{M} = \sqrt{2} \approx 1.4$. For the Bernoulli and Rademacher sensing matrices, the results for equal magnitudes differ from those for other choices of $\mathcal{X}$ reflecting the dependency of $f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X})$ on $\mathcal{X}$. For the Rademacher $\boldsymbol{A}$ the coefficient of variation for both Gaussian and Uniform input signal is somewhat smaller than for the special case of equal (non-zero) signal magnitudes. The shape of the curve however remains largely unchanged. The Bernoulli $\boldsymbol{A}$, on the other hand, demonstrates a different trend. For lower values of K the curves for the generic and special cases coincide. However, with growing K they quickly depart from each other with the former one starting to slightly increase whereas the latter decreasing according to the analytic prediction. Nevertheless, they provide a lower spread of the output SNR than that of a Gaussian $\boldsymbol{A}$. Finally, in all considered cases the coefficient of variation is inversely proportional to M leading to higher values for higher compression rates.

3.3.3 Non-canonical extensions

Having investigated the spread of the output SNR in compressed measurements based on the canonical CS model of (2.22), we now briefly examine how this analysis can be extended to several notable non-canonical model variations that are commonly found in practice; namely, a complex-valued model and a composite model.

3.3.3.1 Complex model

Consider a complex-valued version of (2.24) where $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$ and $\boldsymbol{A} \in \mathbb{C}^{M\times N}$, while $\boldsymbol{n}_{\mathrm{m}} \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{n}_{\mathrm{s}} \in \mathbb{C}^{N\times 1}$ are circularly-symmetric complex noise vectors with entries $n_{\mathrm{m},m} \sim \mathcal{CN}(0, \sigma_{\mathrm{m}})$ and $n_{\mathrm{s},n} \sim \mathcal{CN}(0, \sigma_{\mathrm{s}}^2)$, respectively[12]. Then, (3.13) becomes

$$\eta_{\mathrm{O}} = \frac{1}{M\sigma_0^2} \sum_{m=1}^{M} \left| \sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k} x_k \right|^2 = \frac{1}{M\sigma_0^2} \sum_{m=1}^{M} |d_m|^2. \tag{3.39}$$

Comparing (3.39) with (3.13), we see that the OSNR now depends on $\mathcal{S}(\boldsymbol{x})$ via $|d_m|^2 = |\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k|^2$ where each $a_{m,k}x_k \in \mathbb{C}$.

Gaussian $\boldsymbol{A}$. When the elements of $\boldsymbol{A}$ are drawn from a zero-mean circularly-symmetric complex normal distribution $\mathcal{CN}(0, 1/M)$, d_m is a zero-mean complex normal variable with variance $\sigma_m^2 = \sum_{k=1}^{K} |x_k|^2/M = \|\boldsymbol{x}\|_2^2/M$. It follows then that $|d_m|$ is Rayleigh distributed

[11] The interval $[a, b]$ is chosen such that $\mathrm{var}\{x_i\} = \frac{1}{12}(b-a)^2 = 1/K$.

[12] The notation $n \sim \mathcal{CN}(c, \sigma^2)$ implies that $\mathrm{Re}\{n\} \sim \mathcal{N}(c, \sigma^2/2)$, $\mathrm{Im}\{n\} \sim \mathcal{N}(c, \sigma^2/2)$, and $\mathrm{Re}\{n\}$, $\mathrm{Im}\{n\}$ are independent from each other. For further details see Appendix E.1.

with a scale parameter $\sigma_m/\sqrt{2}$ and therefore, for a given $\mathcal{X}$, $\beta = \sum_{m=1}^{M} |d_m|^2$ is distributed according to the Gamma distribution with the shape parameter M and the scale parameter σ_m^2, i.e.,

$$f_{\boldsymbol{A}}(\beta|\mathcal{X}) = \Gamma\left(M, \frac{\|\boldsymbol{x}\|_2^2}{M}\right). \tag{3.40}$$

From (3.40), it immediately follows that

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}|\mathcal{X}) = \Gamma\left(M, \frac{\vartheta}{M}\right) \tag{3.41}$$

Expressions (3.40)-(3.41) are identical to the ones derived for the real-valued case, which means that all the results obtained above hold for the complex Gaussian $\boldsymbol{A}$ as well.

Bernoulli and Rademacher $\boldsymbol{A}$. Since there is no meaningful complex analogs to Bernoulli and Rademacher sensing matrices, we suppose that $\boldsymbol{A}$ in this case are the same real-valued matrices as before, but the input and the measured signals are complex-valued. Thus, $d_m = \sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k \in \mathbb{C}$. Following the same line of reasoning as in the real-valued case, we constrain ourselves, for the sake of tractability, to the case of all x_k being equal to some (unknown) complex value x. Then, $d_m = x\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}$, where $\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}$ is a real-valued random variable. Taking into account that $|d_m|^2 = |x|^2\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}\right)^2$ and $\sum_{k=1}^{K}|x_k|^2 = K|x|^2$, one can easily see that all results for Bernoulli and Rademacher sensing matrices obtained for the real-valued case hold in the complex case as well.

3.3.3.2 Composite model

The second non-canonical model we look at is the composite extension of (2.22) where the sensing matrix is expressed as the product of two distinct matrices: the measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N\times M}$ and a (non-identity) sparsifying basis $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$. We can write it as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{s} + \boldsymbol{n}, \tag{3.42}$$

where $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$ and $\boldsymbol{s} = \boldsymbol{\Psi}\boldsymbol{x}$. In contrast to (2.22) where we freely design the sensing matrix $\boldsymbol{A}$, here we choose the elements of the measurement matrix $\boldsymbol{\Phi}$ instead.

Given[13] $\boldsymbol{\Psi}$, we can express the elements of $\boldsymbol{A}$ as $a_{m,n} = \boldsymbol{\phi}_m^{\mathrm{T}}\boldsymbol{\psi}_n$ where $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$ denote ith columns of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, respectively. When the elements of the measurement matrix $\boldsymbol{\Phi}$ are drawn from some random distribution, we can first determine the distribution of the elements of $\boldsymbol{A}$ and then proceed to analyze the OSNR spread in the same manner as before. In Appendix B.1.6, we show that for all three matrix types considered so far we arrive at a Gaussian (approximately in case of Bernoulli and Rademacher $\boldsymbol{\Phi}$) sensing matrix $\boldsymbol{A}$ whose elements, however, can possibly have non-zero means and different variances.

[13]Note that the sparsifying basis $\boldsymbol{\Psi}$ is normally defined by the considered application and it is either known a prior or learned from the measurements as in [123], [124], [170] for instance.

Selection $\boldsymbol{\Phi}$. In the context of the composite model, another common type of sensing matrix is a (random) selection matrix. It consists of a selection of M rows of the $N \times N$ identity matrix $\mathbf{I}_N$ such that each row and column of $\boldsymbol{\Phi}$ have at most a single non-zero, respectively. Denoting by n_m the index of the non-zero element in mth column of $\boldsymbol{\Phi}$, we can express the elements of $\boldsymbol{A}$ simply as $a_{m,n} = \psi_{n_m,n}$. In other words, the sensing matrix $\boldsymbol{A}$ constitutes a selection of M distinct rows of $\boldsymbol{\Psi}$.

Compare now the total effective signal power in compressed measurements in this case

$$\beta = \sum_{m=1}^{M} \left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} a_{m,k} x_k \right)^2 = \sum_{m=1}^{M} \left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} \psi_{n_m,k} x_k \right)^2 \tag{3.43}$$

with the total power of the Nyquist rate signal $\boldsymbol{s} = \boldsymbol{\Psi}\boldsymbol{x}$

$$\beta_{\text{Nyq}} = \|\boldsymbol{\Psi}\boldsymbol{x}\|_2^2 = \left\| \sum_{k \in \mathcal{S}(\boldsymbol{x})} \boldsymbol{\psi}_k x_k \right\|_2^2 = \sum_{n=1}^{N} \left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} \psi_{n,k} x_k \right)^2 . \tag{3.44}$$

From (3.43) and (3.44), we see that the selection $\boldsymbol{\Phi}$ can result, depending on $\boldsymbol{\Psi}$, not only in the spread of the effective signal power, but, by sub-sampling $\boldsymbol{s}$, in a reduction of the total signal power inversely proportionally to the compression ratio $\rho = M/N$.

Example 3.4

In this example, we investigate the mean and variance of β/β_{Nyq} for three types of measurement matrices $\boldsymbol{\Phi}$ and two bases $\boldsymbol{\Psi}$: a discrete cosine transform (DCT) and a discrete Fourier transform (DFT) that we denote by $\boldsymbol{\Psi}_{\text{dct}}$ and $\boldsymbol{\Psi}_{\text{dft}}$, respectively. Note that in the latter case, the elements of $\boldsymbol{x}$ are complex-valued. According to our expectations, we see that Gaussian and Rademacher $\boldsymbol{\Phi}$ on average provide similar power (and accordingly SNR) spread with expected signal power equal to that of the Nyquist

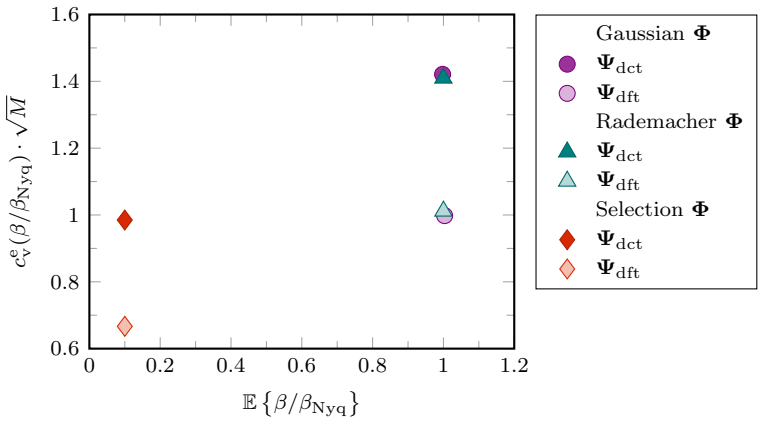


Figure 3.6: Average coefficient of variation $c_{\text{v}}^{\text{e}}(\beta/\beta_{\text{Nyq}})$ vs. the expected value $\mathbb{E}\{\beta/\beta_{\text{Nyq}}\}$ for a measurement matrix $\boldsymbol{\Phi}$ with $N = 300$, $\rho = 0.1$ and $K = 3$.

sampled signal equivalent. The selection measurement matrix, on the other hand, has a lower spread at the expense of the significantly lower expected signal power.

3.4 Impact on recovery performance

Intuitively, it is clear that with regard to the SNR in CS system, one generally wishes that its expected value is close to that of the Nyquist rate system and that it is stable with respect to (non-power related) signal parameters. However, what is of ultimate interest is the recovery performance. Therefore, the two questions we attempt to answer in this section are i) how does the OSNR spread affect the recovery, and ii) how can one account for it during the system design. We begin with the former one and investigate the influence of the OSNR spread on the recovery.

3.4.1 CRB-type performance

Sparse signal recovery can ultimately be viewed as an estimation task, whether the sensing matrix $\boldsymbol{A}$ corresponds to a parametric dictionary or not. A standard way of measuring the performance of an estimator is by comparing it against the Cramér-Rao bound (CRB) that presents the lower bound attainable by any estimation algorithm in a given setting [171]. Having in mind the noisy CS model (2.22), one can show that, in the case of Gaussian noise, the unbiased CRB[14] corresponds to the MSE of the oracle-assisted estimator introduced in Section 3.3.1 [158]. Hence, by evaluating the MSE ϵ_{orcl} given by

$$\epsilon_{\text{orcl}} = \mathbb{E}\left\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2\right\} = \mathbb{E}\left\{\|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{n}\|_2^2\right\} \tag{3.45}$$

we can obtain a CRB-type performance metric. Note that $\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}$ in (3.45) can be written as

$$\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger} = \left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\mathrm{T}}\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}\right)^{-1}\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\mathrm{T}}, \tag{3.46}$$

and, for a fixed $\boldsymbol{A}$, it generally depends on $\mathcal{S}(\boldsymbol{x})$. We express this dependency by writing $\epsilon_{\text{orcl}}(\mathcal{S}(\boldsymbol{x}))$, while understanding the expectation in (3.45) to be taken over the ensemble of $\boldsymbol{n}$.

Oracle MSE for 1-sparse input signals. Following our analysis approach we begin with the special case of $(K = 1)$-sparse signals. The submatrix $\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}$ in this case becomes a column vector $\boldsymbol{a}_n$ where the index n corresponds to the position of the non-zero element in $\boldsymbol{x}$. Substituting $\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})} = \boldsymbol{a}_n$ into (3.46), we have that

$$\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger} = \boldsymbol{a}_n^{\dagger} = \frac{1}{\|\boldsymbol{a}_n\|_2^2}\boldsymbol{a}_n^{\mathrm{T}} = \tilde{\boldsymbol{a}}_n^{\mathrm{T}}, \tag{3.47}$$

[14] The attainability of the CRB in CS setting has been studied in such works as [172] and [173].

where by $\tilde{\boldsymbol{a}}_n$ we denote an nth column of $\boldsymbol{A}$ normalized to its squared ℓ_2-norm. When $\boldsymbol{n}$ is a zero-mean normal Gaussian vector (the matrix $\boldsymbol{A}$ has orthogonal rows of equal norm c), $\tilde{\boldsymbol{a}}_n^{\mathrm{T}}\boldsymbol{n}$ is a zero-mean Gaussian random variable with variance σ_0^2. The MSE ϵ_{orcl} is then equal to $\sigma_0^2 = c\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{m}}^2$ irrespective of the signal position n and the choice of the sensing matrix $\boldsymbol{A}$. Otherwise,

$$\epsilon_{\mathrm{orcl}}(n) = \mathrm{trace}\left(\tilde{\boldsymbol{a}}^{\mathrm{T}}\boldsymbol{\Sigma}\tilde{\boldsymbol{a}}_n\right) = \sigma_{\mathrm{s}}^2\tilde{\boldsymbol{a}}_n^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\tilde{\boldsymbol{a}}_n + \sigma_{\mathrm{m}}^2, \tag{3.48}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{n}$ defined in (3.1).

Oracle MSE for arbitrary sparse input signals. For arbitrary sparse signals, $\epsilon_{\mathrm{orcl}} = \mathbb{E}\left\{\|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{n}\|_2^2\right\}$ can be written as

$$\begin{aligned}\epsilon_{\mathrm{orcl}}(\mathcal{S}(\boldsymbol{x})) &= \mathrm{trace}\left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{\Sigma}\left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\right)^{\mathrm{T}}\right)\\ &= \sigma_{\mathrm{s}}^2\mathrm{trace}\left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\dagger}\right)^{\mathrm{T}}\right) + \sigma_{\mathrm{m}}^2\mathrm{trace}\left(\left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\mathrm{T}}\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}\right)^{-1}\right).\end{aligned} \tag{3.49}$$

In case of orthogonal $\boldsymbol{A}$, (3.49) turns into $\epsilon_{\mathrm{orcl}}(\mathcal{S}(\boldsymbol{x})) = \sigma_0^2\mathrm{trace}\left(\left(\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}^{\mathrm{T}}\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}\right)^{-1}\right)$, which corresponds to the CRB derived in [158].

Example 3.5

To illustrate how the oracle MSE varies depending on the index of the non-zero entry of $\boldsymbol{x}$, Figure 3.7 shows $\epsilon_{\mathrm{orcl}}/M$ as well as its empirical PDF $f(\epsilon_{\mathrm{orcl}}/M)$ in case of $K = 1$ over the index n for a fixed realization of a Gaussian $\boldsymbol{A}$ with $N = 300$, $\rho = 0.1$ and $\sigma_{\mathrm{s}}^2 = \sigma_{\mathrm{m}}^2$. We see that the MSE changes significantly with the peak to average ratio of ≈ 2.9. Ideally, we would like $\mathbb{E}\{\epsilon_{\mathrm{orcl}}\}$ to be minimized with the peak ratio being close to 1. For comparison, Figure (3.8) also demonstrates empirical PDFs of the normalized

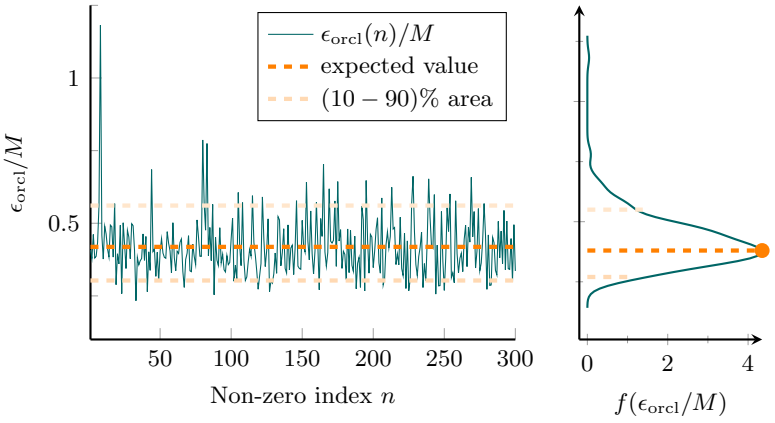


Figure 3.7: Normalized MSE and its empirical PDF of a $K = 1$ sparse signal as a function of the index n for the oracle assisted estimator and a Gaussian $\boldsymbol{A}$ with $N = 300$, $\rho = 0.1$ and $\sigma_{\mathrm{s}}^2 = \sigma_{\mathrm{m}}^2 = 1$.

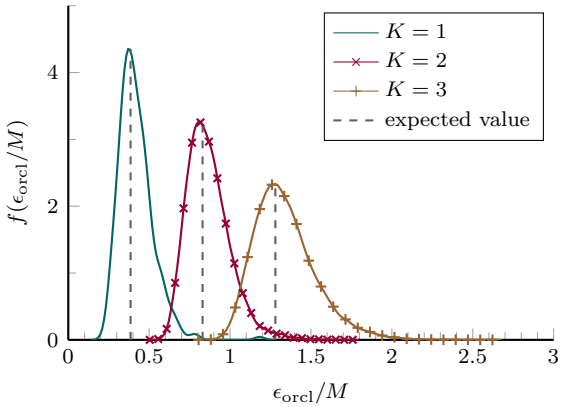


Figure 3.8: Empirical PDF of a normalized MSE of a K-sparse signal over the support $\mathcal{S}(\boldsymbol{x})$ for the oracle assisted estimator and a Gaussian $\boldsymbol{A}$ with $N = 300$, $\rho = 0.1$ and $\sigma_{\text{s}}^2 = \sigma_{\text{m}}^2 = 1$.

oracle MSE ϵ_{orcl}/M for different values of K. We observe that in all cases the oracle MSE exhibits significant spread over the support of the input signal, which in turn means non-uniform CRB and subsequently non-uniform best-case performance. This confirms our analysis of the recovered SNR in Section 3.3.1, where we saw that the spread of the output SNR can potentially lead to support-dependent recovery guarantees.

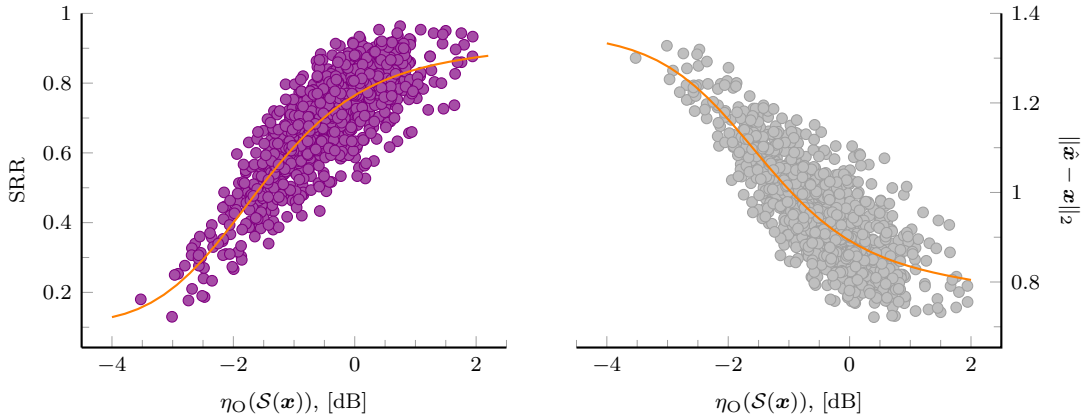
3.4.2 Support recovery performance

Having discussed the CRB-type performance of an oracle-assisted estimator that knows the support of the input signal, we move to analyzing the more challenging part of sparse recovery that is the support estimation[15]. In many of the available efficient recovery algorithms, it involves a basic step of computing a correlation between the measurement data and the dictionary. With this in mind, we focus our analysis on correlation-based approaches.

3.4.2.1 Numerical evaluation

We begin with a numerical investigation of the influence of the OSNR spread on the support recovery performance. To do so, we calculate the empirical support recovery rate (SRR) (the percentage of (partially) correctly estimated supports) and the RMSE between the true ($\boldsymbol{x}$) and the estimated ($\hat{\boldsymbol{x}}$) input signals for signal supports with different values of η_{O} and a fixed random Gaussian $\boldsymbol{A}$. As an underlying recovery method we use the orthogonal matching pursuit (OMP), the greedy algorithm described in Section 2.3.2 that is widely used due to its low computational complexity and satisfactory performance under a range of practical settings. In order to exclude possible effects of the varying dynamic range, we make the magnitudes of all non-zero elements equal to $\frac{1}{\sqrt{K}}$, while setting the rest of the system parameters as follows: $N = 300$, $\rho = 0.2$, $K = 3$, $\varsigma = 2$, and $\eta_{\text{I}} = 20$ dB. The

[15]Note that when the support $\mathcal{S}(\boldsymbol{x})$ is known the estimation of the input signal $\boldsymbol{x}$ becomes trivial, e.g., it can be done by solving a standard LS problem as in (2.29).



(a) Support recovery rate **(b)** RMSE

Figure 3.9: Recovery performance of a Gaussian $\boldsymbol{A}$ as a function of the support $\mathcal{S}(\boldsymbol{x})$ plotted against the corresponding OSNR $\eta_{\mathrm{O}}(\mathcal{S}(\boldsymbol{x}))$: the support recovery rate (a) and the RMSE (b).

results are presented in Figure 3.9, where Figure 3.9a) shows the average (among 10^3 noise realizations) SRR, whereas Figure 3.9b) demonstrates the average RMSE. Both are plotted against the OSNR where the notation $\eta_{\mathrm{O}}(\mathcal{S}(\boldsymbol{x}))$ means that each particular η_{O} corresponds to a different support $\mathcal{S}(\boldsymbol{x})$ according to (3.6). Additionally, we highlight the average trend in both graphs by a solid orange-colored curve. From Figure 3.9 we see that in the resulting OSNR range of $[-4, 2]$ dB, the recovery performance is significantly non-uniform over the support of the input signal. As expected, it (on average) decreases for supports with lower OSNR and increases for those with higher ones. Moreover, a significant part of possible signal supports have the effective SNR below the level needed for reliable (e.g., SRR > 0.8) support recovery with a notable number of them falling below the 0.5 mark. This indicates that the OSNR spread can indeed have a significant impact on the overall recovery performance, especially in the "border" SNR region where a transition to stable recovery happens. To further investigate this effect, in the following we analyze the probability of wrong support estimation on an example of the canonical CS model.

3.4.2.2 Analysis of the error probability

Consider the (canonical) SMV model (2.22) where $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$. For the sake of brevity, we suppose henceforth that $\boldsymbol{n}$ is a white Gaussian noise with variance σ_0^2. Given (2.22), we would like to compute the probability of estimating the wrong support, i.e., $\Pr\left[\hat{\mathcal{S}}(\boldsymbol{x}) \neq \mathcal{S}(\boldsymbol{x})\right]$.

1-sparse input signals. For the special case of $(K = 1)$-sparse signals, we have that

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{a}_i x_i + \boldsymbol{n}, \tag{3.50}$$

where $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$ and i is the index of the (single) non-zero element in $\boldsymbol{x}$. For a correlation-based estimator, the support recovery simplifies to a following single-step

estimation:

$$\hat{\mathcal{S}}(\boldsymbol{x}) = \arg(\max_n |\boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{y}|). \tag{3.51}$$

Given (3.51), we would like to calculate the probability that the estimated support is an index $j \neq i$. This occurs when the correlation of $\boldsymbol{y}$ with $\boldsymbol{a}_j$ is higher than with $\boldsymbol{a}_i$, i.e.,

$$P_{\mathrm{E}}(j|i) = \Pr\left[\hat{\mathcal{S}}(\boldsymbol{x}) = \{j\}|\mathcal{S}(\boldsymbol{x}) = \{i\}\right] = \Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right]. \tag{3.52}$$

Theorem 2. *Let $\gamma_{k,\ell}$ be the inner product between kth and ℓth columns of $\boldsymbol{A}$ such that $\gamma_{k,\ell} = \boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{a}_\ell \in \mathbb{R}$ and $\bar{\eta}_{\mathrm{O}} = \frac{\|\boldsymbol{a}_i\|_2^2 x_i^2}{\sigma_0^2}$ denote the normalized OSNR (i.e., $\bar{\eta}_{\mathrm{O}} = M\eta_{\mathrm{O}}$). Then,*

$$Pr\left[||\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] = Q\left(\sqrt{\bar{\eta}_{\mathrm{O}}}\beta_{j,i}^{-}\right) \cdot Q\left(-\sqrt{\bar{\eta}_{\mathrm{O}}}\beta_{j,i}^{+}\right) + Q\left(-\sqrt{\bar{\eta}_{\mathrm{O}}}\beta_{j,i}^{-}\right) \cdot Q\left(\sqrt{\bar{\eta}_{\mathrm{O}}}\beta_{j,i}^{+}\right), \tag{3.53}$$

where $\beta_{j,i}^{\mp} = \sqrt{\frac{1 \mp \frac{\gamma_{j,i}}{\gamma_{i,i}}}{1 + \frac{\gamma_{j,j} \mp \gamma_{j,i}}{\gamma_{i,i} \mp \gamma_{j,i}}}}$ while $Q(\cdot)$ denotes the Q-function (see App. D).

Proof. Cf. Appendix B.1.7. ∎

Theorem 2 states that the probability of confusing the ith index with the jth, which we denote as $P_{\mathrm{E}}(j|i)$, depends on the square root of the (normalized) OSNR and the column correlations $\gamma_{j,i}, \gamma_{i,i}, \gamma_{j,j}$ via $\beta_{j,i}^{\mp}$. To calculate the total probability of error for ith signal position, one has to consider maximum correlation among all possible $j \neq i$, i.e.,

$$P_{\mathrm{E}}(i) = \Pr\left[\hat{\mathcal{S}}(\boldsymbol{x}) \neq \mathcal{S}(\boldsymbol{x})|\mathcal{S}(\boldsymbol{x}) = \{i\}\right] = \Pr\left[\max_j |\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right], \tag{3.54}$$

which makes it difficult to analyze analytically. However, we can bound (3.54) as

$$P_{\mathrm{E}}(j_{\mathrm{worst}}|i) \leq P_{\mathrm{E}}(i) \leq P_{\mathrm{U}}(i) = \sum_{j \neq i} P_{\mathrm{E}}(j|i). \tag{3.55}$$

The upper bound in (3.55) is a union bound [174] on (3.54), whereas the lower bound in (3.55) is given by considering the pair (i, j_{worst}) that corresponds to the highest correlation $|\gamma_{k,i}|$ for a given index i, i.e., $j_{\mathrm{worst}} = \arg(\max_j |\gamma_{j,i}|)$. Note that $P_{\mathrm{E}}(j_{\mathrm{worst}}|i) \leq P_{\mathrm{E}}(j_{\mathrm{worst}}|i_{\mathrm{worst}})$ where the index pair $(i_{\mathrm{worst}}, j_{\mathrm{worst}})$ corresponds to the coherence of $\boldsymbol{A}$.

Example 3.6

To verify our derivations, we compare the empirical probability of erroneous support estimation $P_{\mathrm{E}}^{\mathrm{emp}}(i)$ with the analytic bounds (3.55) for a given index i and a single realization of a Gaussian $\boldsymbol{A}$ with $N = 300$ and $\rho = 0.1$. Figure 3.10 demonstrates the empirical error probability $P_{\mathrm{E}}^{\mathrm{emp}}(i)$ calculated over 10^5 noise realizations as a function of

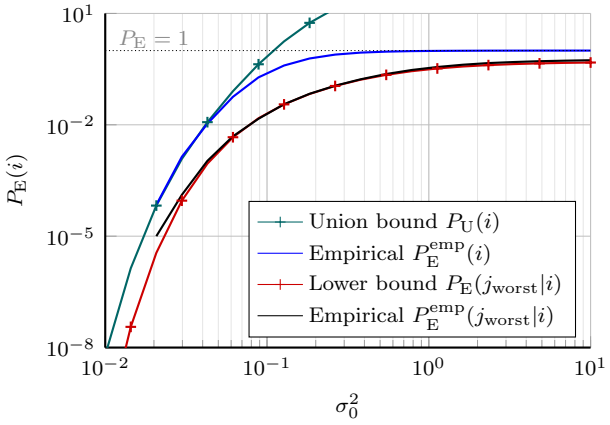
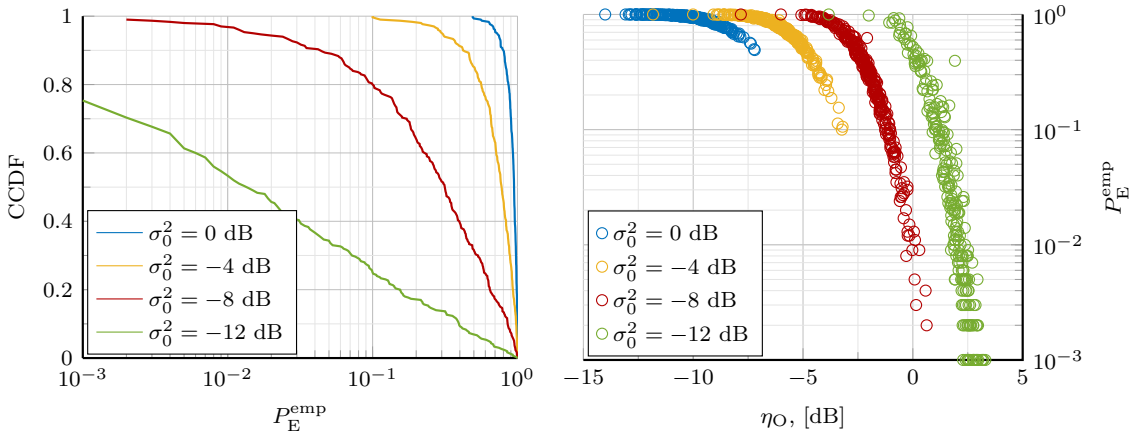


Figure 3.10: Probability of erroneous support estimation $P_{\mathrm{E}}(i)$ for a 1-sparse $\boldsymbol{x}$ of unit power and a single realization of a Gaussian $\boldsymbol{A}$ with $N = 300$ and $\rho = 0.1$.

the noise power σ_0^2 where $x_i = 1$ as well as the analytic upper and lower bounds provided by $P_{\mathrm{U}}(i)$ and $P_{\mathrm{E}}(j_{\mathrm{worst}}|i)$, respectively. For comparison, we also plot the empirical probability $P_{\mathrm{E}}^{\mathrm{emp}}(j_{\mathrm{worst}}|i)$, which coincides with the analytic curve since (3.53) is exact. Contrarily, the union bound provides a close approximation of $P_{\mathrm{E}}(i)$ in the low noise regime, while starting to diverge from it with the increase of σ_0^2. The difference between the lower bound $P_{\mathrm{E}}(j_{\mathrm{worst}}|i)$ and $P_{\mathrm{E}}^{\mathrm{emp}}(i)$, on the other hand, remains largely constant.

Note that $P_{\mathrm{E}}(j|i)$ and therefore $P_{\mathrm{E}}(i)$ depend on the position of the non-zero element in $\boldsymbol{x}$ via $\beta_{j,i}^{\mp}$ and $\bar{\eta}_{\mathrm{O}}(i) = \frac{\gamma_{i,i}x_i^2}{\sigma_0^2}$. To evaluate the variations of $P_{\mathrm{E}}(i)$ over the signal support, we plot in Figure 3.11a the CCDF of $P_{\mathrm{E}}^{\mathrm{emp}}(i)$ for a scenario from Example (3.6) with several values of σ_0^2 (note that $x_i = 1\ \forall i$). It shows that the error probability experiences large variations over the signal position, whose range increases with the decrease of the noise power



(a) CCDF of $P_{\mathrm{E}}^{\mathrm{emp}}$

(b) $P_{\mathrm{E}}^{\mathrm{emp}}$ vs. $\eta_{\mathrm{O}} = \bar{\eta}_{\mathrm{O}}/M$

Figure 3.11: Influence of the OSNR spread on P_{E}: complimentary cumulative distribution function (CCDF) of $P_{\mathrm{E}}^{\mathrm{emp}}$ over the non-zero index i (a) and $P_{\mathrm{E}}^{\mathrm{emp}}$ vs. η_{O} (b) for a 1-sparse $\boldsymbol{x}$ of unit power and a single realization of a Gaussian $\boldsymbol{A}$ with $N = 300$ and $\rho = 0.1$.

(i.e., increase of the SNR). To isolate the influence of the support-related OSNR spread on P_{E}, Figure 3.11b presents $P_{\mathrm{E}}^{\mathrm{emp}}(i)$ vs. corresponding $\eta_{\mathrm{O}}(i)$ for a fixed value of noise power σ_0^2. Note that different values of $\eta_{\mathrm{O}}(i)$ for a given σ_0^2 in this case result from the change of the support (i.e., the index i) only. We observe a clear relation between the spread of the OSNR and the spread of P_{E}: the larger is the OSNR for a given noise and signal power, the smaller is $P_{\mathrm{E}}(i)$ and hence the better is the support recovery performance. This corresponds to the results presented in Figure 3.9 confirming that the OSNR spread negatively affects the probability of correct support recovery.

Arbitrary sparse input signals. In case of ($K > 1$)-sparse input signals (3.50) becomes

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} = \sum_{i \in \mathcal{S}(\boldsymbol{x})} \boldsymbol{a}_i x_i + \boldsymbol{n}. \tag{3.56}$$

For an iterative correlation-based recovery algorithm such as the OMP for instance, we can define the probability of estimating a wrong non-zero index $j \notin \mathcal{S}(\boldsymbol{x})$ at the first iteration:

$$\begin{aligned} P_{\mathrm{E}}(j|\mathcal{S}(\boldsymbol{x})) &= \Pr\left[\hat{\mathcal{S}}(\boldsymbol{x}) = \{j\} | j \notin \mathcal{S}(\boldsymbol{x})\right] = \Pr\left[\bigcap_{i \in \mathcal{S}(\boldsymbol{x})} \left(|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right)\right] \\ &= \Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq \max_i |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right]. \end{aligned} \tag{3.57}$$

The "max" condition in (3.57) complicates the calculation of $P_{\mathrm{E}}(j|\mathcal{S}(\boldsymbol{x}))$. However, we can lower-bound (3.57) by

$$P_{\mathrm{E}}(j|i_k) < P_{\mathrm{E}}(j|\mathcal{S}(\boldsymbol{x})), \tag{3.58}$$

where $P_{\mathrm{E}}(j|i_k) = \Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_{i_k}^{\mathrm{T}}\boldsymbol{y}|\right]$ and $i_k = \arg\left(\max_i |\gamma_{i,i}|\right) \in \mathcal{S}(\boldsymbol{x})$.

Theorem 3. *Let $\gamma_{k,\ell}$ be the inner product between kth and ℓth columns of $\boldsymbol{A}$ such that $\gamma_{k,\ell} = \boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{a}_\ell \in \mathbb{R}$. Then, for an arbitrary sparse signal we have that*

$$Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] = Q\left(\alpha_{j,i}^{-}\right) \cdot Q\left(-\alpha_{j,i}^{+}\right) + Q\left(-\alpha_{j,i}^{-}\right) \cdot Q\left(\alpha_{j,i}^{+}\right), \tag{3.59}$$

where $\alpha_{i,j}^{\mp} = \sqrt{\frac{\left(\sum_{k \in \mathcal{S}(\boldsymbol{x})} (\gamma_{i,k} \mp \gamma_{j,k}) x_k\right)^2}{(\gamma_{i,i} + \gamma_{j,j} \mp 2\gamma_{j,i})\sigma_0^2}}$.

Proof. Cf. Appendix B.1.8. ■

Note that since $\bar{\eta}_{\mathrm{O}} = \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\sigma_0^2} = \frac{\|\sum_m \sum_{k \in \mathcal{S}(\boldsymbol{x})} \boldsymbol{a}_k x_k\|_2^2}{\sigma_0^2}$, in the general case we cannot represent the Q-function argument $\alpha_{j,i}^{\mp}$ via $\bar{\eta}_{\mathrm{O}}$ directly as in (3.53). We can do so however for the special case of $x_k = x\ \forall k$ according to the following corollary.

Corollary 1. *Let* $\bar{\eta}_{\mathrm{O}}^{(i)} = \frac{\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}x^2}{\sigma_0^2}$ *denote a partial normalized output SNR where* $\bar{\eta}_{\mathrm{O}} = \sum_{i\in\mathcal{S}(\boldsymbol{x})}\bar{\eta}_{\mathrm{O}}^{(i)}$. *Then, if* $x_k = x\ \forall k \in [1, K]$, *the probability* $Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right]$ *is given by*

$$Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] = Q\left(\sqrt{\bar{\eta}_{\mathrm{O}}^{(i)}}\beta_{j,i}^{-}\right)\cdot Q\left(-\sqrt{\bar{\eta}_{\mathrm{O}}^{(i)}}\beta_{j,i}^{+}\right) + Q\left(-\sqrt{\bar{\eta}_{\mathrm{O}}^{(i)}}\beta_{j,i}^{-}\right)\cdot Q\left(\sqrt{\bar{\eta}_{\mathrm{O}}^{(i)}}\beta_{j,i}^{+}\right), \tag{3.60}$$

where $\beta_{i,j}^{\mp} = \sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)\right)^2}{\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}\cdot(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})}}$.

Proof. Cf. Appendix B.1.9. ■

Corollary 1 shows how the individual probabilities $\Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right]$ depend on the partial OSNRs $\bar{\eta}_{\mathrm{O}}^{(i)}$ corresponding to the ith signal entry.

Applying Theorem 3 to (3.58), we arrive at the lower bound for $P_{\mathrm{E}}(j|\mathcal{S}(\boldsymbol{x}))$

$$P_{\mathrm{E}}(j, i_k) = Q\left(\alpha_{j,i_k}^{-}\right)\cdot Q\left(-\alpha_{j,i_k}^{+}\right) + Q\left(-\alpha_{j,i_k}^{-}\right)\cdot Q\left(\alpha_{j,i_k}^{+}\right) < P_{\mathrm{E}}(j|\mathcal{S}(\boldsymbol{x})), \tag{3.61}$$

where $i_k = \arg\left(\max_i|\gamma_{j,i}|\right) \in \mathcal{S}(\boldsymbol{x})$ and

$$\left(\alpha_{i,j}^{\mp}\right)^2 = \begin{cases} \frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)\right)^2}{\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}\cdot(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})}\bar{\eta}_{\mathrm{O}}^{(i)}, & \text{if } x_k = k\ \forall k \\ \frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)x_k\right)^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}, & \text{otherwise} \end{cases}.$$

Given (3.61) and (3.58), for the total error probability $P_{\mathrm{E}}(\mathcal{S}(\boldsymbol{x})) = \Pr\left[\hat{\mathcal{S}}(\boldsymbol{x}) = \{j\} \not\subset \mathcal{S}(\boldsymbol{x})\right]$ we can write that

$$P_{\mathrm{E}}(j_{\mathrm{worst}}, i_k) < P_{\mathrm{E}}(\mathcal{S}(\boldsymbol{x})), \tag{3.62}$$

where j_{worst} corresponds to an index $j \notin \mathcal{S}(\boldsymbol{x})$ that yields the largest correlation $|\gamma_{j,i}| : i \in \mathcal{S}(\boldsymbol{x})$, i.e., $j_{\mathrm{worst}} = \arg\left(\max_{i\in\mathcal{S}(\boldsymbol{x}),j}|\gamma_{j,i}|\right)$.

Example 3.7

Figure 3.12 depicts the empirical probability of erroneous support estimation $P_{\mathrm{E}}^{\mathrm{emp}}(\mathcal{S}(\boldsymbol{x}))$ calculated over 10^5 noise realizations together with the analytic lower bound $P_{\mathrm{E}}(j_{\mathrm{worst}}, i_k)$ for a given support set $\mathcal{S}(\boldsymbol{x})$ of different size and a single realization of a Gaussian $\boldsymbol{A}$ with $N = 300$ and $\rho = 0.1$. Note that the presented results correspond to a single fixed support set chosen such that $\mathcal{S}(\boldsymbol{x})_{|K=n} \subset \mathcal{S}(\boldsymbol{x})_{|K=n+1}, n = 1, 2, 3$, with the magnitudes of the non-zero elements being $x_i = \sqrt{\frac{1}{K}}\ \forall i \in \mathcal{S}(\boldsymbol{x})$. We see that although the lower bound $P_{\mathrm{E}}(j_{\mathrm{worst}}, i_k)$ becomes less tight with the increase of K, it largely follows the overall trend: the probability of correct estimation decreases with growing K.

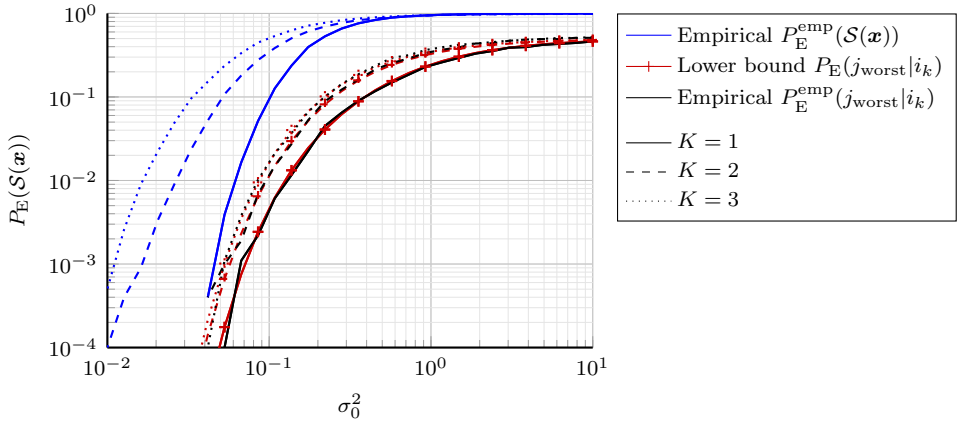
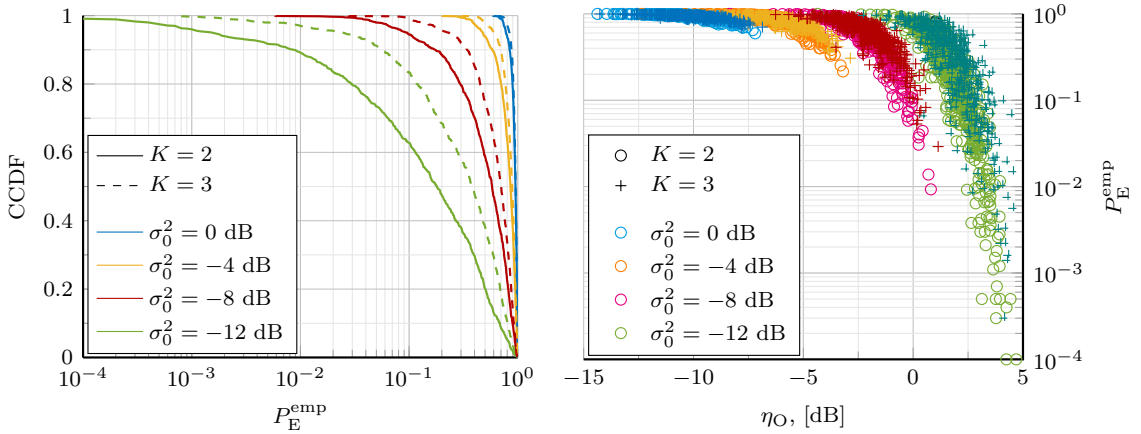


Figure 3.12: Probability of erroneous support estimation $P_{\mathrm{E}}(\mathcal{S}(\boldsymbol{x}))$ for a K-sparse $\boldsymbol{x}$ of unit power, i.e., $\|\boldsymbol{x}\|_2^2 = 1$, and a single realization of a Gaussian $\boldsymbol{A}$ with $N = 300$ and $\rho = 0.1$.

Finally, in Figure 3.13 we evaluate the variation of the error probability over the support $\mathcal{S}(\boldsymbol{x})$ and its relation to the spread of the OSNR for the setting from Example 3.6. Thus, Figure 3.13a shows the CCDF of $P_{\mathrm{E}}^{\mathrm{emp}}$ over $\mathcal{S}(\boldsymbol{x})$, whereas Figure 3.13b depicts the spread of $P_{\mathrm{E}}^{\mathrm{emp}}$ with respect to the OSNR η_{O}. Note that both figures are obtained by evaluating $P_{\mathrm{E}}(\mathcal{S}(\boldsymbol{x}))$ for a random subset of possible supports of size $L = 500$ since an exhaustive check would require $L = \frac{N!}{K!(N-K)!}$. We observe here a similar tendency as in the 1-sparse case: i) the error probability significantly varies over the support of the input signal while ii) this variation is clearly coupled to the change of the OSNR.



(a) CCDF of $P_{\mathrm{E}}^{\mathrm{emp}}$

(b) $P_{\mathrm{E}}^{\mathrm{emp}}$ vs. η_{O}

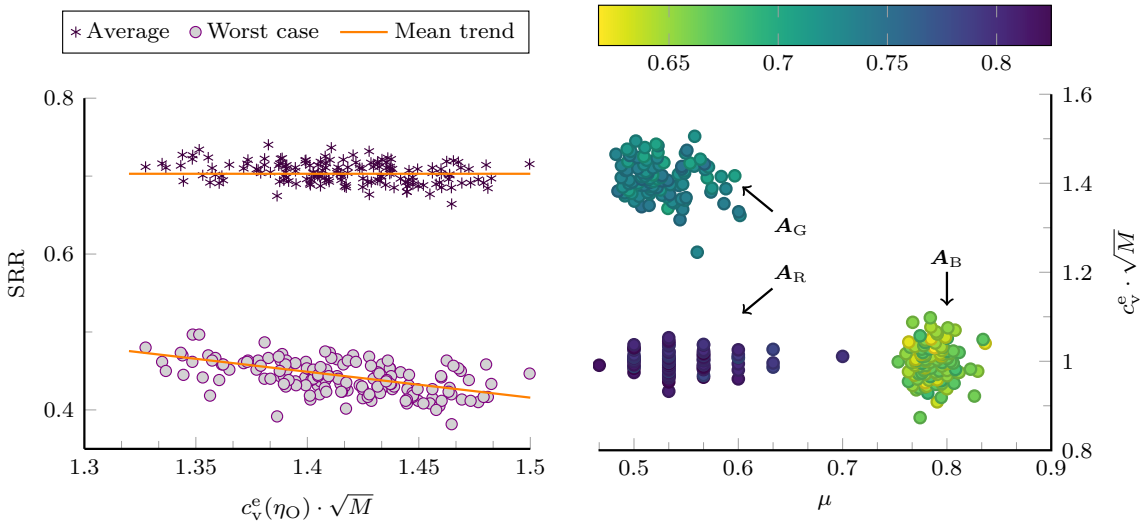
Figure 3.13: Influence of the OSNR spread on P_{E}: CCDF of $P_{\mathrm{E}}^{\mathrm{emp}}$ over the support $\mathcal{S}(\boldsymbol{x})$ (a) and $P_{\mathrm{E}}^{\mathrm{emp}}$ vs. η_{O} (b) for a K-sparse $\boldsymbol{x}$ of unit power and a single realization of a Gaussian $\boldsymbol{A}$ with $N = 300$ and $\rho = 0.1$.

3.5 Some design ideas

This brings us to the second question posed at the beginning of this section: how can one take into account the negative influence of the OSNR spread on the recovery performance during the measurement design? Obviously, the ultimate design goal would be to avoid the spread altogether as it is in the Nyquist setting where $\mathbf{\Phi} = \mathbf{I}_N$ and the OSNR is constant. Our analysis shows however, that in the CS setting this effect is essentially unavoidable. A possible way to cope with it is to design the measurement such as to minimize the variations of the output SNR. This alone, however, cannot guarantee an improvement of the overall system performance because one has to take into account other matrix properties essential for signal recovery, such as the matrix coherence for instance. In the following, we investigate possible ways to approach the measurement matrix design for improved recovery performance.

3.5.1 Random design

We begin by examining random matrix designs as the ones considered at the beginning of this chapter. On one hand, random matrices are known to fulfill a number of theoretical guarantees and provide a *universal* measurement kernel in the sense that they are likely to be incoherent to any input signal. On the other hand, to be implemented in hardware one normally requires to have a fixed measurement kernel rather than a random ensemble. Although, as mentioned in Section 3.3.1, concentration results for random matrices tell us that the performance of the ensemble asymptotically concentrates around some mean [167], which implies that it is irrelevant which particular realization is used, practical experience suggests that for limited dimensions this is often not good enough. One practical solution is to generate a large number of sensing matrix candidates and pick only the "best" ones for practical implementation, e.g. those that minimize the OSNR spread for a certain level of matrix coherence. We evaluate this strategy by conducting numerical experiments. In our first experiment, we pick 100 realizations of a Gaussian sensing matrix that have similar (with the difference on the order of 10^{-2}) coherence but provide different values of the OSNR spread and compute the average and the worst-case (according to the 5% level) support recovery rate (SRR), which we depict in Figure 3.14a. We see that while the average recovery rate stays largely constant at the level corresponding to the expected value of the OSNR, the worst case performance decreases nearly linearly with the increase of the OSNR spread. In the second experiment, we calculate the average SRR for 50 different realizations of Gaussian ($\boldsymbol{A}_{\mathrm{G}}$), Bernoulli ($\boldsymbol{A}_{\mathrm{B}}$) and Rademacher ($\boldsymbol{A}_{\mathrm{R}}$) sensing matrices and depict it in a form of a colored scatter plot in Figure (3.14)b. The best performance unsurprisingly corresponds to the matrices that have both the lower coherence and the lower SNR spread, whereas the matrices that have notably higher coherence on average perform significantly worse. These results demonstrate that within small deviations of the matrix coherence, the strategy of picking the sensing matrix with a smaller OSNR spread pays off (e.g., compare



(a) Gaussian $\boldsymbol{A}$ with $\mu = 0.51$ **(b)** Average SRR

Figure 3.14: Support recovery performance of sensing matrices with $N = 300$, $\rho = 0.2$, $K = 3$, $\varsigma = 2$, and $\eta_{\mathrm{I}} = 20$ dB: a Gaussian $\boldsymbol{A}$ with a fixed μ vs. $c_{\mathrm{v}}^{\mathrm{e}}(\eta_{\mathrm{O}})$ (a) and different $\boldsymbol{A}$ vs. μ and $c_{\mathrm{v}}^{\mathrm{e}}(\eta_{\mathrm{O}})$ (b).

the results for $\boldsymbol{A}_{\mathrm{G}}$ and $\boldsymbol{A}_{\mathrm{R}}$ that have similar coherence values but significantly different SNR spreads). However, choosing the sensing matrix only with respect to the OSNR spread is not a good strategy as it might result in a higher coherence and, subsequently, an overall inferior performance as is evident from Figure (3.14)b.

Alternatively, one can avoid random generation altogether by designing the measurement kernel in a deterministic manner. For instance, it has been proposed to utilize the matrix coherence as an objective function for matrix optimization [156], [175]–[178]. It is also possible to use more application specific performance measures as demonstrated in [A3] that considers measurement matrix design for DoA estimation with compressive antenna arrays. Building on the results achieved in [A3], below we provide an example of how such a design approach can be applied to more generic sparse recovery problems.

3.5.2 Optimized design

3.5.2.1 Design framework

Instead of drawing the elements of the measurement/sensing matrix randomly and hope for a suitable outcome, we formulate the design task as a constrained optimization problem:

$$\boldsymbol{A}_{\mathrm{opt}} = \arg\left(\min_{\boldsymbol{A}} J(\boldsymbol{A})\right) \text{ s.t. } \mathcal{C}(\boldsymbol{A}, \alpha, \beta, \cdots) \tag{3.63}$$

in case of the canonical model (2.23) and

$$\boldsymbol{\Phi}_{\mathrm{opt}} = \arg\left(\min_{\boldsymbol{\Phi}} J(\boldsymbol{\Phi})\right) \text{ s.t. } \mathcal{C}(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \alpha, \beta, \cdots) \tag{3.64}$$

in case of the composite one (3.42). In both (3.63) and (3.64), $J(\cdot)$ denotes an objective function defined by the application and $\mathcal{C}(\cdots)$ represents a set of optimization constraints. Note that (3.63) corresponds to the case when $\boldsymbol{\Psi} = \mathbf{I}_N$ and $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Psi} = \boldsymbol{\Phi}^{\mathrm{H}}$. In the following, we propose two particular formulations of (3.63)–(3.64): a design based on the probability of erroneous support estimation and a correlation-based design. In the former, we utilize lower bounds on the error probability derived at the end of the previous section. In the latter, we design the measurement with respect to the structure of the correlations between the columns of the sensing matrix in an attempt to provide improved accuracy in terms of the index distance between true and erroneous estimates.

3.5.2.2 Design examples

Probability-based design. In our first design example, we formulate (3.64) in terms of the probability of wrong support estimation for a certain SNR level with a constraint on the OSNR spread. To do so, we use the lower bound (3.62) on the error probability derived in Section 3.4.2.2 which yields the following cost function

$$J\left(\boldsymbol{\Phi}\right) = \max_{i,j|\vartheta_{\mathrm{E}}} P_{\mathrm{E}}(j,i), \tag{3.65}$$

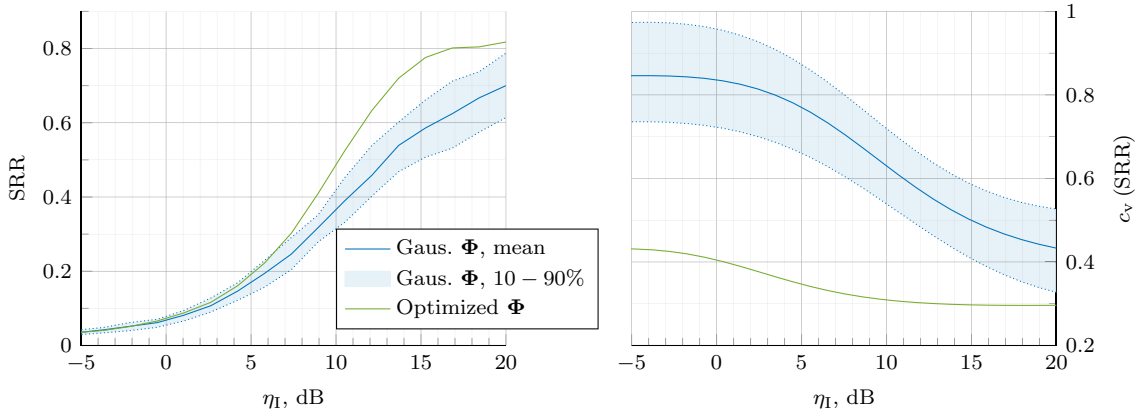
where $P_{\mathrm{E}}(j,i) = \Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right]$ is given by (3.60) and $\vartheta_{\mathrm{E}} = \frac{P_{\mathrm{s}}}{M\sigma_0^2}$ is the (average) input SNR. Here, the notation $\max\limits_{a|b}$ means that the search over a variable a is performed for a given value of b. Using (3.65), we complete (3.64) as

$$\boldsymbol{\Phi}_{\mathrm{opt}} = \arg\left(\min_{\boldsymbol{\Phi}} \max_{i,j|\vartheta_{\mathrm{E}}} P_{\mathrm{E}}(j,i)\right) \text{ s.t. } \begin{cases} c_{\mathrm{v}}^{\mathrm{emp}}(\tilde{\beta}) \leq \zeta \\ \mu(\boldsymbol{A}) \leq \eta \end{cases}, \tag{3.66}$$

where $c_{\mathrm{v}}^{\mathrm{emp}}(\tilde{\beta})$ is the empirical coefficient of variation of $\tilde{\beta} = \|\boldsymbol{A}_{\mathcal{S}(\boldsymbol{x})}\|_2^2$ over the support $\mathcal{S}(\boldsymbol{x}) : |\mathcal{S}(\boldsymbol{x})| = K$, whereas ζ and η are thresholds we are free to choose. Note that, according to the discussion in Section 3.3.1, the OSNR spread is (largely) defined by $\beta = \|\boldsymbol{A}\boldsymbol{x}\|_{\mathrm{F}}^2$. Constraining $c_{\mathrm{v}}^{\mathrm{emp}}(\beta)$, however, requires searching over the set of input signals $\boldsymbol{x}$, which is impractical both from the viewpoint of the computational complexity and the applicability of the results. On the other hand, limiting $c_{\mathrm{v}}^{\mathrm{emp}}(\tilde{\beta})$ provides a means of controlling the variation of the OSNR, although indirectly, while being independent of the input signal $\boldsymbol{x}$. The constraint on the sensing matrix coherence $\mu(\boldsymbol{A}) \leq \eta$ is needed to ensure compatibility with the CS recoverability requirements.

Example 3.8

We evaluate the design strategy (3.66) for a following choice of parameters: $N = 50$, $M = 10$, $K = 1$ and $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Psi}$ where $\boldsymbol{\Psi}$ is an $N \times N$ DCT matrix. To solve (3.66), we



(a) Support recovery rate

(b) Coefficient of variation of SRR

Figure 3.15: Comparison of SRR (a) and its coefficient of variation (b) for a Gaussian $\mathbf{\Phi}$ and a $\mathbf{\Phi}$ optimized according to (3.66) where $N = 50$, $M = 10$, and $\mathbf{\Psi}$ is an $N \times N$ DCT matrix.

use the fmincon function[a] of the MATLAB numerical optimization toolbox [179] where the optimization parameters are chosen as follows: $\zeta = 1/\sqrt{M}$, $\eta = 0.87$, and $\sigma_0 = 10^{-2}$ where $\vartheta_{\mathrm{E}} = 1/M\sigma_0^2$. We access the performance of the support recovery by computing the SRR and calculating its coefficient of variation $c_{\mathrm{v}}(\mathrm{SRR})$ over the support of the input signal which we show in Figure (3.15)a and b, respectively. Along with the results for a measurement matrix $\mathbf{\Phi}$ obtained by solving (3.66) we also present the mean and $(10-90)$-percentile area for 50 realizations of a Gaussian $\mathbf{\Phi}$. We observe that in terms of the SRR the optimized measurement matrix $\mathbf{\Phi}$ allows to achieve a SRR similar to the one provided by the (best) random Gaussian matrix, while being significantly superior with respect to its coefficient of variation over the signal's support. In other words, our results indicate that $\mathbf{\Phi}$ optimized according to (3.66) yields lower performance variations over the support of the input signal than the random Gaussian matrices and hence provides more stable average-to-worst case performance.

[a]Note that (3.66) is a highly non-convex problem. As such, an application of numerical optimization routines to solving it is likely to result in finding one of the local minima. To avoid sub-optimal solutions, one can perform several optimization trials with (different) random function initialization. Our numerical evidence suggests that in the case of (3.66) these, however, produce solutions that are only negligibly different in terms of the cost function (3.65).

Correlation-based design. Here we provide an example of a slightly different perspective on the measurement matrix design. As briefly discussed in Section 2.2.1, the sparse recovery framework can also be applied to solving a parameter estimation problem, rather than a signal reconstruction one. The support of the sparse input in this case can be interpreted as defining a parameter set in the parameter space represented by the basis $\mathbf{\Psi}$. The distance between the true and the estimated signal supports determines then the accuracy of the parameter estimation. To illustrate this, consider two estimated parameter sets, e.g., obtained

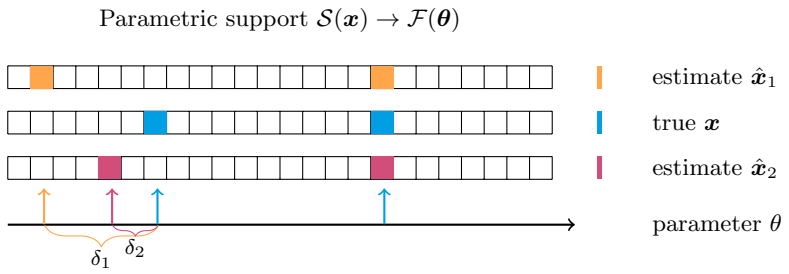


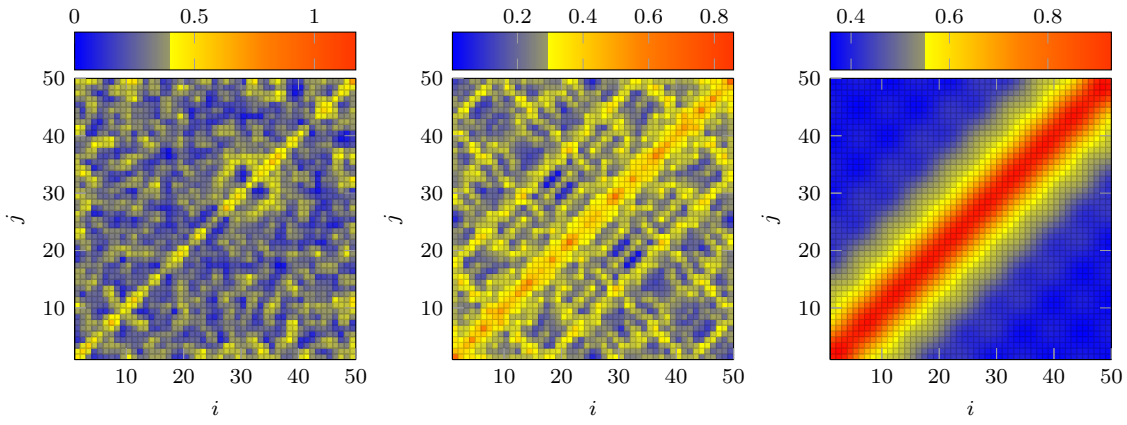
Figure 3.16: The difference between $\mathcal{S}(\boldsymbol{x})$ and $\mathcal{S}(\hat{\boldsymbol{x}})$ determines the distance between the true $\boldsymbol{x}$ and the estimated $\hat{\boldsymbol{x}}$ parameter vectors: a closer estimate $\hat{\boldsymbol{x}}_2$ with the (parameter) distance δ_2 from $\boldsymbol{x}$ is preferable to the farther estimate $\hat{\boldsymbol{x}}_1$ with the distance $\delta_1 > \delta_2$.

by applying two different estimation algorithms: one is closer to the true parameter set in the parameter space, whereas the other one is farther away from it as exemplified in Figure 3.16. While both estimates might have the same ℓ_0, ℓ_1 and even ℓ_2 distance from the true set, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}_1\|_p = \|\boldsymbol{x} - \hat{\boldsymbol{x}}_2\|_p$ for $p = 0, 1$, and 2, from the parameter estimation viewpoint the former is preferable to the latter.

The idea behind measurement matrix optimization for parameter estimation is to increase the probability of obtaining an erroneous estimate that is closer to the true parameter in the parameter space. In previous sections we saw that, given a fixed overall SNR, the main factor defining the probability of a wrong estimate for correlation based recovery algorithms is the column correlations $\gamma_{i,j} = \boldsymbol{a}_i^{\mathrm{H}} \boldsymbol{a}_j$ of the sensing matrix $\boldsymbol{A}$. Having this in mind, we base our design on the Gram matrix $\boldsymbol{G}_{\boldsymbol{A}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Psi}$ whose (i,j)th element is $\gamma_{i,j}$.

Example 3.9

In this example, we examine the structure of the Gram matrix for three different choices of the sensing matrix: a (real) random Gaussian $\boldsymbol{A}$ ($\boldsymbol{A}_{\mathrm{G}}$), a Vandermonde structured $\boldsymbol{A}$ obtained by randomly selecting M rows of an $N \times N$ DCT matrix ($\boldsymbol{A}_{\mathrm{V}}$), and a structured $\boldsymbol{A}$



(a) Gaussian $\boldsymbol{A}$ **(b)** Vandermonde $\boldsymbol{A}$ **(c)** Toeplitz $\boldsymbol{A}$

Figure 3.17: Gram matrices $|\boldsymbol{G}_{\boldsymbol{A}}|$ for three types of sensing matrix with $N = 50$ and $M = 10$: random Gaussian $\boldsymbol{A}$ (a), Vandermonde $\boldsymbol{A}$ (b), and Toeplitz $\boldsymbol{A}$ (c).

($\boldsymbol{A}_{\mathrm{T}}$) obtained by truncating the singular-value decomposition (SVD) of an $N \times N$ square symmetric Toeplitz matrix with a first column $\boldsymbol{t}_1 = [1, e^{1-\beta}, e^{1-2\beta}, \ldots, e^{1-(N-1)\beta}]^{\mathrm{T}}$ where $\beta = \frac{1}{N^5}$ (i.e., $\boldsymbol{A}_{\mathrm{T}} = [\boldsymbol{S}]_{1:M}^{1/2}\boldsymbol{U}^{\mathrm{T}}$ where $\boldsymbol{T} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^{\mathrm{T}}$ and the (diagonal) elements of $\boldsymbol{S}$ are assumed to be arranged in a descending order). The Gram matrices of all three of them are illustrated in Figure 3.17 for the case of $N = 50$ and $M = 10$. We notice a difference in the level of structure present in the Gramian: from no particular structure for $\boldsymbol{A}_{\mathrm{G}}$ to some degree of concentration of the (off-diagonal) elements with higher magnitude around the main diagonal for $\boldsymbol{A}_{\mathrm{V}}$, to their exponential decay for $\boldsymbol{A}_{\mathrm{T}}$.

In Example 3.9, we observe that random design yields a non-structured $\boldsymbol{G}_{\boldsymbol{A}}$ with elements whose magnitude is independent of their index distance to the main diagonal . In case of the wrong support recovery, this is likely to result in an arbitrarily large index distance between the erroneous and true estimate and, subsequently, an arbitrarily large parameter estimation error[16] that is irrespective of the SNR, which contradicts our design goals.

Example 3.10

To investigate how the different structure of the Gram matrix affects the support estimation accuracy we evaluate the index distance between the true and estimated supports by means of the so-called earth mover's distance (EMD) (also known as 1st Mallows or 1st Wasserstein distance) [180]. Originally introduced as a metric for comparison of distributions [180]–[182], the EMD turned out to be useful as a feature similarity measure, particularly in image analysis [183]. It has also been used as an objective function in the CS context as an alternative to ℓ_0/ℓ_1-norm minimization [50], [184], [185]. In [A9], we show that the EMD provides an effective metric for evaluating support distance, a property that we will use here. More details on the definition and calculation of EMD can be found in Appendix A.

Figure 3.18 demonstrates the average EMD as a function of the ISNR for $N = 50$ and $M = 10$. It shows the average EMD calculated for the entire support, i.e., $d_{\mathrm{EMD}}^{\mathrm{T}} = d_{\mathrm{EMD}}\left(\mathcal{S}(\boldsymbol{x}), \mathcal{S}(\hat{\boldsymbol{x}})\right)$ where $d_{\mathrm{EMD}}(\mathcal{A}, \mathcal{B})$ denotes the EMD between two index sets $\mathcal{A}, \mathcal{B}$, and for the erroneously estimated support part only, i.e., $d_{\mathrm{EMD}}^{\mathrm{E}} = d_{\mathrm{EMD}}\left(\mathcal{S}(\hat{\boldsymbol{x}})/\mathcal{S}(\boldsymbol{x}), \mathcal{S}(\boldsymbol{x})/\mathcal{S}(\hat{\boldsymbol{x}})\right)$ where $\mathcal{A}/\mathcal{B}$ denotes the relative complement of $\mathcal{B}$ in $\mathcal{A}$. We see that the sensing matrix with the Toeplitz structured Gramian consistently provides a better support estimation accuracy, as judged by the index distance. The less structured Gaussian and Vandermonde matrices on the other hand yield lower support estimation accuracy in terms of the EMD. Note that since $d_{\mathrm{EMD}}^{\mathrm{E}}$ accounts only for the distance of the wrongly estimated part of the support, it does not reflect overall increase in the support recovery rate. On the contrary, $d_{\mathrm{EMD}}^{\mathrm{T}}$ takes into account both. Therefore,

[16] Note that this is not an issue from the CS recovery perspective since traditional CS recovery metrics, such as the MSE or SRR, do not differentiate between a "closer" and a "farther" error [A9].

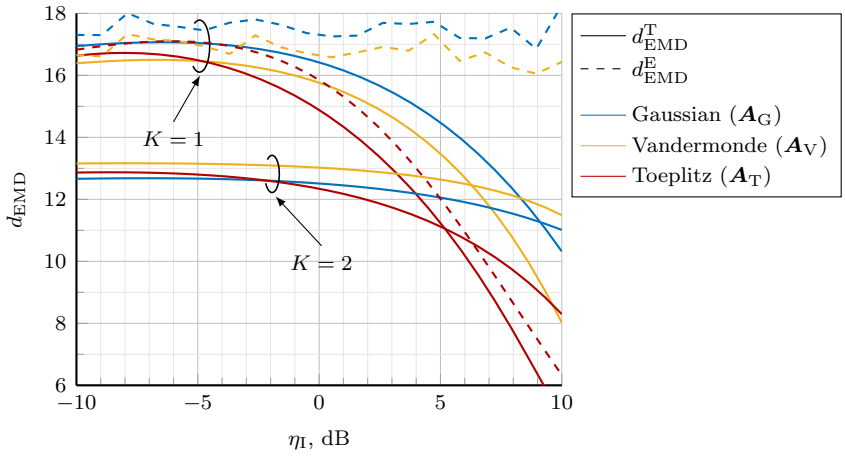


Figure 3.18: Average EMD for three types of sensing matrix (Gaussian ($\boldsymbol{A}_{\text{G}}$), Vandermonde ($\boldsymbol{A}_{\text{V}}$), and Toeplitz ($\boldsymbol{A}_{\text{T}}$)) with $N = 50$ and $M = 10$. The EMD is lower bounded by 0 and upper bounded by $N - K$.

$d^{\text{T}}_{\text{EMD}}$ decreases for all three choices of $\boldsymbol{A}$ as the SNR increases, while $d^{\text{E}}_{\text{EMD}}$ improves only for structured $\boldsymbol{A}_{\text{T}}$ staying largely constant for other choices of the sensing matrix. On the other hand, for structured sensing matrices we expect to have higher column correlations and, subsequently, higher matrix coherence, which is known to negatively impact the CS recovery performance, e.g., expressed by the MSE. We will come back to this later when examining the proposed design strategy. Note that in our example here we have $\mu(\boldsymbol{A}_{\text{G}}) = 0.85 < \mu(\boldsymbol{A}_{\text{V}}) = 0.91 < \mu(\boldsymbol{A}_{\text{T}}) = 0.99$.

Ideally, we would like $\boldsymbol{G}_{\boldsymbol{A}}$ to be an identity or at least a diagonal matrix, implying the orthogonality of the columns of $\boldsymbol{A}$, which is impossible to achieve since for $M < N$ we cannot have more than M mutually orthogonal columns. Several works have thus focused on optimizing $\boldsymbol{A}$ such that its Gramian provides a good approximation (e.g., in terms of the Frobenius norm) to $\mathbf{I}_N$ [175], [176]. Instead, here we aim at the design where $\boldsymbol{G}_{\boldsymbol{A}}$ has a quasi-decaying structure, i.e., the elements with a higher magnitude are likely to appear close to the main diagonal whereas the ones with the lower magnitudes concentrate around the corners, e.g., as in the Toeplitz matrix from Figure 3.17c. In example 3.9 we saw, however, that designing $\boldsymbol{A}$ only with respect to the structure of its Gramian results in a matrix whose "standard" CS properties, such as the coherence for instance, might deteriorate. Therefore, one needs to introduce additional constraints that will allow a compromise between the desired level of structure in $\boldsymbol{G}_{\boldsymbol{A}}$ and the CS properties of $\boldsymbol{A}$.

To enforce a decaying structure into $\boldsymbol{G}_{\boldsymbol{A}}$ we formulate the objective function in (3.64) as

$$J(\boldsymbol{\Phi}) = \|\boldsymbol{G}_{\boldsymbol{A}} \odot \boldsymbol{T}\|_{\text{F}}, \tag{3.67}$$

where $\boldsymbol{T}$ is an $N \times N$ square symmetric Toeplitz matrix whose first column is $\boldsymbol{t}_1 = [0, 1, \ldots, N-1]^{\text{T}}$ and $\odot$ denotes the Hadamard (element-wise) matrix multiplication. The

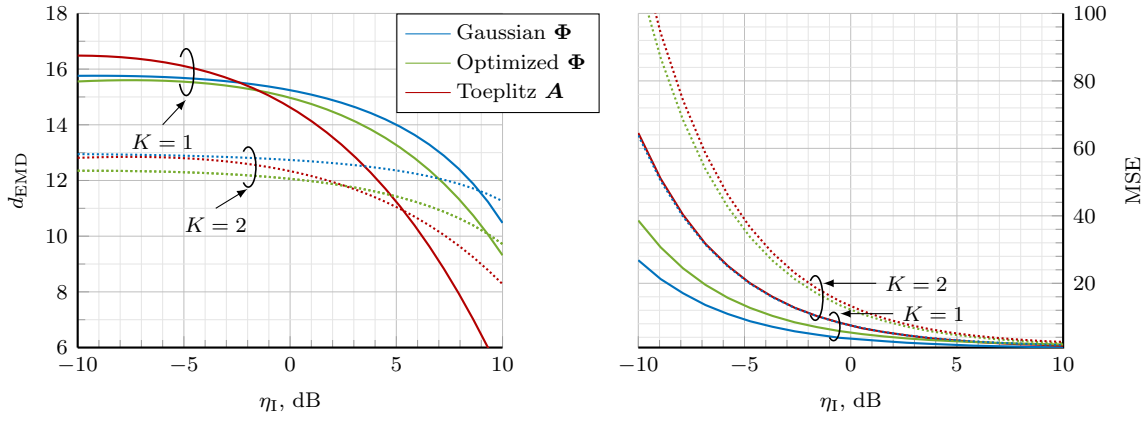
matrix $\boldsymbol{T}$ in (3.67) weights the elements of $\boldsymbol{G_A}$ proportional to their distance from the main diagonal: the bigger is the distance the larger is the weight. To avoid trivial solutions and ensure fulfillment of the CS recovery requirements we constrain the elements of the main diagonal $\boldsymbol{d} = \operatorname{diag}(\boldsymbol{G_A}) \in \mathbb{R}^{N\times 1}$ of $\boldsymbol{G_A}$ and the coherence μ of $\boldsymbol{A}$ such that

$$\boldsymbol{\Phi}_{\text{opt}} = \arg\left(\min_{\boldsymbol{\Phi}} \|\boldsymbol{G_A} \odot \boldsymbol{T}\|_{\text{F}}\right) \text{ s.t. } \begin{cases} \zeta_{\text{l}} \leq d_n \leq \zeta_{\text{u}} \; \forall n = 1, \ldots, N \\ \mu(\boldsymbol{A}) \leq \eta \end{cases}, \tag{3.68}$$

where d_n is the nth element of $\boldsymbol{d}$. The bounds $\zeta_{\text{l}}, \zeta_{\text{u}}$, and η should be chosen according to the particular problem formulation and its dimensions.

Example 3.11

In the following example, we apply the optimization strategy (3.68) to the scenario from Example 3.8 where we additionally introduce an $N \times N$ DCT basis $\boldsymbol{\Psi}$ such that $\boldsymbol{A} = \boldsymbol{\Phi}^{\text{T}}\boldsymbol{\Psi}$. Note that although (3.68) is convex with respect to $\boldsymbol{G_A}$, constraining its rank such that $\operatorname{rank}(\boldsymbol{G_A}) \leq \operatorname{rank}(\boldsymbol{\Phi}) \leq M < N$ makes solving (3.68) non-trivial. Therefore, instead of attempting to find a convex formulation of (3.68), we solve it numerically by applying the fmincon function of MATLAB, as in Example 3.8. The constraints on d_n and $\mu(\boldsymbol{A})$ are chosen according to the corresponding average values for a sensing matrix with Gaussian $\boldsymbol{\Phi}$, namely $\zeta_{\text{l}} = 0.5$, $\zeta_{\text{u}} = 1.5$, and $\eta = 0.93$. We evaluate the resulting sparse recovery performance by means of EMD and MSE in Figures 3.19a and b, respectively. For reference, we also present the results for the Gaussian $\boldsymbol{\Phi}$ and the Toeplitz sensing matrix. The graphs show that the optimized design provides an improved index distance accuracy compared to the unstructured Gaussian matrix, while expectedly performing worse than it with respect to the overall MSE due to the large



(a) EMD **(b)** MSE

Figure 3.19: Comparison of an average EMD (a) and an MSE (b) for a Gaussian $\boldsymbol{\Phi}$, an $\boldsymbol{\Phi}$ optimized according to (3.68), and a Toeplitz $\boldsymbol{A}$ where $N = 50$, $M = 10$, and $\boldsymbol{\Psi}$ is an $N \times N$ DCT matrix. The corresponding matrix coherence values are 0.83, 0.91, and 0.99, respectively.

difference in coherence (recall the remark at the end of Example 3.10). On the other hand, compared to the Toeplitz matrix the trend is reversed, the optimized $\boldsymbol{\Phi}$ is better with respect to MSE but provides a worse EMD. This outlines the trade-off between the recovery performance in terms of the Euclidean and index distances. Note that we can influence the balance between the two even further by introducing an additional weighting matrix in (3.67), as explained in [A3, Section III-B] for instance.

3.6 Concluding remarks

In this chapter, we studied the influence of the sensing matrix $\boldsymbol{A}$ in the setting of the canonical CS model (2.22) (or the measurement matrix $\boldsymbol{\Phi}$ in the composite model (3.42)) on the system performance. We began with considering noise and signal parameters at the system's output, namely the characteristics of the white input noise and the effective SNR. We demonstrated that, in contrast to Nyquist rate sampling, the application of the CS framework leads to variations of the output SNR (OSNR) over the support of the input signal. As a result, the achievable performance also becomes support-dependent, an effect that is often overlooked during CS performance evaluation. To evaluate the range of such variations, we investigated the OSNR spread for three types of sensing matrices common in CS, i.e., random Gaussian, Bernoulli and Rademacher matrices. Our analytic analysis supported by the numerical study indicates that the OSNR spread becomes more pronounced for higher compression rates (lower number of measurements M) and can vary depending on the choice of $\boldsymbol{A}$.

Although the dependency of the effective SNR on the non-power related signal parameters is an interesting phenomenon by itself, its main importance lies in the way it affects the recovery performance. In light of that, in the second half of this chapter we studied the CS performance of the commonly used CS matrices with respect to the change of the support. We started with looking at the CRB-type performance defined by the MSE of the oracle-assisted estimator and then proceeded to investigating the support recovery performance for the case of correlation-based estimators such as the OMP. We saw that, irrespective of the metric used, the recovery performance varies significantly depending on the support of the input signal. Moreover, these variations clearly correspond to the variations of the effective SNR: the supports that yield lower OSNRs on average perform worse than those that yield higher ones. This supports our intuition that the spread of the OSNR negatively affects the overall system performance, especially with respect to the relation between the average and best/worst case estimation accuracy. Our analysis also suggests that random measurement matrix designs so ubiquitously advocated in CS, might not be the best choice, especially for practical applications where parameter settings are far from asymptotic.

Having established the connection between the OSNR spread and the CS recovery performance, we concluded our investigations by outlining some alternative sensing/measurement

matrix design strategies that could allow controlling such system parameters as the degree of OSNR variations. Particularly, we presented two examples of a performance-driven design based on a generic design framework that attempts to optimize the measurement matrix according to the specific requirements at hand. In the first one, we aimed at a sensing matrix that would help mitigating the effect of the OSNR spread, while in the second one we looked at the measurement matrix design from a parameter estimation viewpoint. In both cases, we observed an improved performance with respect to the performance metric of choice. The main drawback of such an optimized design approach is that the associated optimization problem is unlikely to admit a closed-form solution in general. However, it can be usually solved by numerical optimization routines which is an acceptable solution in many practical scenarios when the measurement kernel is not required to change too often. Finally, by adopting a more application-oriented approach to measurement matrix design we lose the advantage of the universality of the measurement kernel enjoyed by random matrices. This could restrict the applicability of this method, e.g., to the cases when the sparsifying basis is fixed and (largely) known.

Chapter 4

Sparsity order estimation

4.1 Motivation and related work

As discussed in previous chapters, the ability to represent a signal as sparse is a key concept in the CS framework. We saw that there exists a vast amount of theoretical results showing under which conditions the recovery of sparse signals can be achieved efficiently, and most of them are expressed in terms of the signal's *sparsity order*, i.e., the number of non-zero coefficients K in a sparsity promoting representation. Obviously, the sparsity order has a tremendous impact on the recovery stage, in particular it determines how many measurements are required for successful recovery. Moreover, many efficient reconstruction algorithms benefit from the knowledge of the signal sparsity (as for instance do OMP, CoSaMP and LASSO). Particularly, selecting an appropriate stopping criterion is a well-known problem for greedy recovery algorithms [186], [187]. When the sparsity order is known, it can be used to help in the assessment of the quality of the current solution and hence in determining the correct termination moment. However, there exist a large number of applications where the sparsity order is not known beforehand and may even vary with time. In such cases, it would be desirable if we could estimate it from the measurement data. It would be even better if we could do it before we run the recovery algorithm, using an estimator that is either less complex than the reconstruction itself and/or more robust to noise. This would potentially allow adapting the reconstruction strategy, i.e., to choose a recovery algorithm whose performance and complexity is best suited to the current sparsity order. Also, it could provide a feedback to the measurement stage to adapt the measurement effort to the complexity of the current signal (scene), provided that the application allows for it.

One way to obtain a sparsity order estimate is to use a cross-validation approach as in [188], [189]. Cross-validation is a technique widely used in statistics and learning theory in which a data set is separated into two: a training and a test set. The estimation results based on the former are then validated over the latter to prevent model overfitting. Unfortunately, this requires performing multiple signal reconstructions, at a significant cost in terms of the computational complexity. In [190], the authors estimate a lower bound on the sparsity level

given by the so-called numerical sparsity, i.e., a ratio between ℓ_1- and ℓ_2- signal norms, from the measurement data itself. To calculate the numerical sparsity, they use a technique known as sketching in which an ℓ_p-norm of a vector is estimated through a number of randomized measurements [191]. It relies on the properties of the distribution from which the elements of the measurement matrix are drawn and generally requires an application of several different types of random matrices. This underlines the main drawback of this approach as it necessitates a repeated change of the sensing kernel whose choice is restricted by the sparsity order estimation (SOE) needs (particularly, the authors in [190] use Cauchy and Gaussian random matrices). In [192], the authors also rely on the properties of the measurement matrix. However, they consider a sparse sensing kernel and calculate a probability of obtaining a zero measurement and formulate a maximum likelihood (ML) type of sparsity order estimator. This method however, is likely to be strongly susceptible to noise as it relies on the presence of exactly zero entries in the measurement vector. Furthermore, sparse matrices are known to have higher coherence compared to their non-sparse counterparts.

In this chapter, we investigate an alternative SOE approach that operates directly in the compressed domain, without the need to perform multiple signal reconstructions or sensing matrix changes. It builds upon the ideas from principal component analysis and parameter estimation with model order selection (MOS). We begin with considering a full-rank multiple measurement vector (MMV) model where the compressed observations represent a linear mixture superimposed by additive noise so that the sparsity order is equal to the effective rank of the observation matrix such that usual MOS techniques can be applied for its estimation. Under a strong limitation on the sample size, the performance of the available MOS algorithms depends on the knowledge of the noise model and may deteriorate significantly when the actual noise statistics are different. To account for this, we propose a method that explicitly considers the measurement process by exploiting an empirical distribution of the noise eigenvalues obtained during a training/calibration period, i.e., when only the receiver noise is present. Hence, it goes by the name of empirical threshold test (ETT). In ETT, one can influence the trade-off between the probability of missed detection (order underestimation) and false alarm (order overestimation) by selecting a detection threshold value. This is particularly important in the CS setting as underestimation might result in the recovery performance deterioration when the estimated order is used to assist the recovery and/or adapt the number of measurements. Interestingly, a close relation between sparse signal reconstruction and parameter estimation with MOS has been already realized in [193]. Another closely related discussion can be found in [194], where the authors consider the CS framework for DoA estimation. However, neither of these works focuses on the SOE problem per se.

Note that the MMV signal model that allows for SOE is based on the availability of multiple snapshots of the mixture of signals and the fact that the signals are incoherent (which implies that they must change in time). Although, the general CS setup does not impose any restrictions on the signal but its sparsity, there is a number of applications where

these assumptions hold. Examples of such applications include multiband communications, radio localization, radar signal processing, etc. It is worth noting that the problem of sparsity order estimation in the particular context of multiband communications has recently received some attention [195]–[197]. They, however, differ from the method considered here in several important aspects. Thus, some, like [195] for instance, are based on heuristic techniques, e.g., by relying on numerous Monte-Carlo simulations, which one would normally like to avoid in a practical system. Others, such as [196] and [197], do not consider practical sub-Nyquist sampling architectures but rather derive their analysis based on significantly simplified models that cannot be readily extended to a realistic analog sub-Nyquist sampling setting. In contrast, an approach presented in this chapter can be applicable is readily applicable to practical settings, as we demonstrate in the next chapter.

Often, the signal support is not constant during the entire observation time but exhibits variations from one block of data to another. We refer to the MMV with such a varying but quasi-static support quasi-stationary. The ability to detect the presence and the moment of the support change is important for recovery performance, the more so the larger is the difference between the supports. This is because the support change increases the overall sparsity order, challenging the exploitation of the joint sparsity for signal recovery. In the worst case, the total sparsity order can even exceed the maximal value, resulting in the recovery breakdown. A prevalent way to account for the possibly changing support is to incorporate a dynamic update mechanism of one form or another directly into the recovery step. The basic idea is to apply recursive reconstruction that appropriately penalizes the presence of innovations in the newly-arrived measurements compared to already available ones [198]–[200]. In Section 4.3 of this chapter, we propose a different approach to tackling this problem that does not require numerous iterative reconstruction steps. We first detect the moment of the support change by applying the SOE technique described above to the windowed portion of the measurement data, and then perform joint recovery on a sequence of identified static MMVs.

Following the discussion of the stationary and quasi-stationary multiple snapshot case, we approach the problem of sparsity order estimation from a single data snapshot in Section 4.4. Surprisingly, it turns out that we can recover linear independence in a single vector of observations as well, by designing the measurement matrix in a certain way. We analytically show that a specific Khatri-Rao design of the measurement matrix is a necessary and a sufficient condition for this and discuss the choice of the parameters. Further details on suitable matrix constructions for Khatri-Rao design for SOE can be found in [201].

The material of Sections 4.2 and 4.3 of this chapter is partially presented in [A16] and [A14], respectively, while Section 4.4 is largely based on [A18].

4.2 Multiple snapshot sparsity order estimation

Consider the MMV model described by (2.31), i.e.,

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}, \qquad \text{(rep. of (2.31))}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] \in \mathbb{C}^{N \times T}$ is jointly K-sparse and $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]^{\mathrm{H}} \in \mathbb{C}^{M \times T}$ contains a collection of consequent measurements $\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t$ of some K-sparse signals $\boldsymbol{x}_t$ acquired via $\boldsymbol{A} = \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Psi}$. In particular, we focus on the full-rank version of (2.31) in which $T > K$ and the individual signals $\boldsymbol{x}_t$ are non-collinear such that $\operatorname{rank}(\boldsymbol{X}) = K$. Note that in the following we refer to the index t as a snapshot index reflecting the fact that each vector $\boldsymbol{x}_t$ can be viewed as a different measurement of the same signal scene. The non-collinearity assumption implies that the individual snapshots $\boldsymbol{x}_t$ exhibit a meaningful (in terms of their linear independence) change from snapshot to snapshot. An example of signal types that fall into this category include a mixture of independently modulated signals commonly found in wireless communication systems, navigation, radar, and array processing.

4.2.1 Rank preservation in full-rank MMV

Given (2.31), our task is to determine the value of K from the matrix of compressed measurements $\boldsymbol{Y}$ by operating entirely in the compressed domain, i.e., without resorting to estimation of the matrix of input signals $\boldsymbol{X}$ (or any of the individual snapshots $\boldsymbol{x}_t$). In order to do so we exploit the fact that $\operatorname{rank}(\boldsymbol{X}) = K$ and the results of the following theorem that provides conditions for the rank preservation in the setting of (2.31).

Definition 8. *A Kruskal-rank* $\operatorname{krank}(\boldsymbol{A})$ *of a matrix* $\boldsymbol{A}$ *is the maximum number* K *such that any* K *columns of* $\boldsymbol{A}$ *are linearly independent.*

Theorem 4 (Lemma 1 in [202])**.** *Let* $\boldsymbol{X}$ *be a rank-*K *row* K*-sparse* $N \times T$ *matrix with* $T > K$ *and* $\boldsymbol{A}$ *be an* $M \times N$ *matrix with* $K < M < N$*. Then, the rank of* $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$ *is equal to* K *if* $\operatorname{krank}(\boldsymbol{A}) \geq K$*.*

Proof. Cf. Appendix B.2.1 ■

Theorem (4) implies that any Kruskal-rank-K sensing matrix $\boldsymbol{A}$ preserves the rank of $\boldsymbol{X}$. This allows us to estimate the sparsity order K as the effective rank of $\boldsymbol{Y}$, provided that the sensing matrix has a required Kruskal-rank. Interestingly, it turns out that any sensing matrix that complies with the CS requirements in the considered setting would allow for sparsity order estimation as well. To see why, consider the MMV recovery condition (2.34). It states that for a full-rank $\boldsymbol{X}$ the necessary and sufficient condition is given by $\operatorname{spark}(\boldsymbol{A}) > K + 1$. On the other hand, for any $M \times N$ matrix $\boldsymbol{A}$ with $\operatorname{rank}(\boldsymbol{A}) < N$ it is known that $\operatorname{spark}(\boldsymbol{A}) \leq \operatorname{krank}(\boldsymbol{A}) + 1$ [203]. Combining the two, we see that (2.34) implies that $\operatorname{krank}(\boldsymbol{A}) \geq K - 1$, which suffices for the rank preservation in (2.31).

4.2.2 Eigenvalue based sparsity order estimation

Having determined that in the full-rank MMV rank $(\boldsymbol{Y}) = K$, we can apply rank estimation techniques to determine the sparsity order K. The task of rank estimation in linear mixture models is closely related to the model order selection problem typically found in parametric approaches to signal processing [204], [205], as well as the principal component analysis [206].

4.2.2.1 Estimation approach

Consider the noisy CS setting of (2.22) where each measurement vector $\boldsymbol{y}_t$ is given by

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{n}_t. \tag{4.1}$$

As before, the noise vector $\boldsymbol{n}_t$ is composed as $\boldsymbol{n}_t = \boldsymbol{A}\boldsymbol{n}_{\mathrm{st}} + \boldsymbol{n}_{\mathrm{mt}}$, where $\boldsymbol{n}_{\mathrm{st}}$ and $\boldsymbol{n}_{\mathrm{mt}}$ indicate (t)th snapshots of signal and measurement noise, respectively. We can now define an $M \times M$ covariance matrix $\boldsymbol{R}_{\mathrm{y}}$ as

$$\boldsymbol{R}_{\mathrm{y}} = \mathbb{E}\left\{\boldsymbol{y}_t \boldsymbol{y}_t^{\mathrm{H}}\right\}. \tag{4.2}$$

Inserting (4.1) into (4.2), we obtain

$$\boldsymbol{R}_{\mathrm{y}} = \boldsymbol{A}\mathbb{E}\left\{\boldsymbol{x}_t \boldsymbol{x}_t^{\mathrm{H}}\right\}\boldsymbol{A}^{\mathrm{H}} + \mathbb{E}\left\{\boldsymbol{n}_t \boldsymbol{n}_t^{\mathrm{H}}\right\} = \boldsymbol{A}\boldsymbol{R}_{\mathrm{x}}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{\Sigma}, \tag{4.3}$$

with $\boldsymbol{R}_{\mathrm{x}} = \mathbb{E}\left\{\boldsymbol{x}_t \boldsymbol{x}_t^{\mathrm{H}}\right\} \in \mathbb{C}^{N \times N}$ and $\boldsymbol{\Sigma}$ being the noise covariance matrix. Note that for zero-mean white noise vectors $\boldsymbol{n}_{\mathrm{st}} \sim \mathcal{N}\left(0, \sigma_{\mathrm{s}}^2\right)$ and $\boldsymbol{n}_{\mathrm{mt}} \sim \mathcal{N}\left(0, \sigma_{\mathrm{m}}^2\right)$, $\boldsymbol{\Sigma} = \sigma_{\mathrm{s}}^2 \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} + \sigma_{\mathrm{m}}^2 \mathbf{I}_M$.

Let us now have a closer look at $\boldsymbol{A}\boldsymbol{R}_{\mathrm{x}}\boldsymbol{A}^{\mathrm{H}}$. Since the individual snapshots $\boldsymbol{x}_t$ are non-collinear and at most K-sparse, $\boldsymbol{R}_{\mathrm{x}}$ has only K non-zero rows whose indices correspond to $\mathcal{S}(\boldsymbol{X})$. This means that, when factorized according to the eigendecomposition, only K out of M eigenvalues of $\boldsymbol{A}\boldsymbol{R}_{\mathrm{x}}\boldsymbol{A}^{\mathrm{H}}$ are non identically zero. Therefore, under white noise conditions (when both $\boldsymbol{n}_{\mathrm{st}}$ and $\boldsymbol{n}_{\mathrm{mt}}$ are white and the sensing matrix $\boldsymbol{A}$ is composed of orthogonal rows of equal norm), the sparsity order is given by the number of normalized eigenvalues that are greater than 1:

$$\frac{\lambda_m}{\lambda_M} = \begin{cases} \frac{\lambda_{\mathrm{s},m}}{\lambda_M} + 1 & , 1 \leq m \leq K \\ 1 & , K < m \leq M \end{cases}. \tag{4.4}$$

Here $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$ denote the ordered set of eigenvalues of $\boldsymbol{R}_{\mathrm{y}}$ and $\lambda_M = \sigma_0^2$ is the variance of each element of $\boldsymbol{n}_t$, whereas $\lambda_{\mathrm{s},m}$ represents the ordered set of K non-zero eigenvalues of $\boldsymbol{A}\boldsymbol{R}_{\mathrm{x}}\boldsymbol{A}^{\mathrm{H}}$. The values of $\lambda_{\mathrm{s},m}$ in (4.4) depends on both $\boldsymbol{x}_t$ and $\boldsymbol{A}$.

However, as discussed in Section 3.2, for an arbitrary sensing matrix with $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} \neq c\mathbf{I}_M$, the noise $\boldsymbol{n}_t$ ceases to be white[1]. To account for this, we can perform prewhitening to the output vector $\boldsymbol{y}_t$ by multiplying it with $(\boldsymbol{\Sigma})^{-1/2}$. After the prewhitening, our observation

[1]Note that even if $\boldsymbol{A}$ has orthogonal rows by design, its implementation in hardware might result in some deviations from the ideal and, as a result, possible noise coloring.

model (4.1) is transformed into

$$\bar{\boldsymbol{y}}_t = (\boldsymbol{\Sigma})^{-1/2}\boldsymbol{y}_t = (\boldsymbol{\Sigma})^{-1/2}\left(\boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{n}_t\right) = \boldsymbol{C}\boldsymbol{x}_t + \tilde{\boldsymbol{n}}_t, \tag{4.5}$$

where $\bar{\boldsymbol{y}}_t = (\boldsymbol{\Sigma})^{-1/2}\boldsymbol{y}_t$ is a whitened observation vector, while $\boldsymbol{C} = (\boldsymbol{\Sigma})^{-1/2}\boldsymbol{A}$ and $\tilde{\boldsymbol{n}}_t$ is a white noise vector with covariance matrix $\mathbf{I}_M$. The covariance matrix $\boldsymbol{R}_{\bar{\mathrm{y}}}$ of the whitened observations then becomes

$$\boldsymbol{R}_{\bar{\mathrm{y}}} = \boldsymbol{C}\boldsymbol{R}_{\mathrm{x}}\boldsymbol{C}^{\mathrm{H}} + \mathbf{I}_M, \tag{4.6}$$

where $\boldsymbol{R}_{\bar{\mathrm{y}}} = \mathbb{E}\left\{\bar{\boldsymbol{y}}_t\bar{\boldsymbol{y}}_t^{\mathrm{H}}\right\} \in \mathbb{C}^{M\times M}$. At this point, we could apply the sparsity order estimation test based on (4.4) where λ_m are now the eigenvalues of $\boldsymbol{R}_{\bar{\mathrm{y}}}$, while $\lambda_{\mathrm{s},m}$ correspond to the eigenvalues of $\boldsymbol{C}\boldsymbol{R}_{\mathrm{x}}\boldsymbol{C}^{\mathrm{H}}$ and $\lambda_M = 1$.

In practice, instead of the true covariance matrix $\boldsymbol{R}_{\bar{\mathrm{y}}}$ one has access to its estimation $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}}$, calculated from a limited[2] number of snapshots T as

$$\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}} = \frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{y}}_t\bar{\boldsymbol{y}}_t^{\mathrm{H}} = \frac{1}{T}\bar{\boldsymbol{Y}}\bar{\boldsymbol{Y}}^{\mathrm{H}} = \boldsymbol{C}\hat{\boldsymbol{R}}_{\mathrm{x}}\boldsymbol{C}^{\mathrm{H}} + \hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}} + \hat{\boldsymbol{R}}_{\mathrm{x},\tilde{\mathrm{n}}}, \tag{4.7}$$

where $\bar{\boldsymbol{Y}} = [\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_T]$ and $\tilde{\boldsymbol{R}}_{\mathrm{x}} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}$, while $\hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}} = \frac{1}{T}\tilde{\boldsymbol{N}}\tilde{\boldsymbol{N}}^{\mathrm{H}}$ denotes the sample noise covariance matrix where $\tilde{\boldsymbol{n}} = [\tilde{\boldsymbol{n}}_1, \ldots, \tilde{\boldsymbol{n}}_T]$ and $\hat{\boldsymbol{R}}_{\mathrm{x},\tilde{\mathrm{n}}}$ is a cross term defined as

$$\hat{\boldsymbol{R}}_{\mathrm{x},\tilde{\mathrm{n}}} = \frac{1}{T}\Big(\left(\boldsymbol{C}\boldsymbol{X}\right)\tilde{\boldsymbol{N}}^{\mathrm{H}} + \tilde{\boldsymbol{N}}\left(\boldsymbol{C}\boldsymbol{X}\right)^{\mathrm{H}}\Big). \tag{4.8}$$

Let the eigenvalues of the sample covariance matrix $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}}$ be given by $\hat{\lambda}_m$, $m = 1, 2, \ldots, M$. Due to the limited number of observations, $\{\hat{\lambda}_m\}_{m=1}^{M}$ differ significantly from the ideal eigenvalue profile from (4.4). First, the noise eigenvalues $\hat{\lambda}_{\mathrm{n},m}$ are not equal to 1 but vary around it resulting in a decaying profile of the ordered set of eigenvalues (see Figure 4.1a in Example 4.12) [207]. Second, the cross term $\hat{\boldsymbol{R}}_{\mathrm{x},\tilde{\mathrm{n}}}$ between the signal and the noise components becomes non-vanishing which further distorts the eigenvalue profile (see Figure 4.1b in Example 4.12). As a result, the distinction between the first K "signal" eigenvalues[3] and the remaining $M - K$ noise ones becomes less straightforward. This calls for alternative estimation approaches compared to the one in (4.4). At this, point classical MOS algorithms developed in the context of the detection of the number of signals can be applied in order to discriminate between the "signal" and noise eigenvalues and hence determine the sparsity order K. Among these, the best known are the information theoretic based methods, such as the Akaike's information criterion (AIC) [208], the minimum description length (MDL) [209], [210] and the efficient detection criterion (EDC) [211], as well as methods that build on the statistical characteristics of the sample eigenvalues. Examples of the latter include

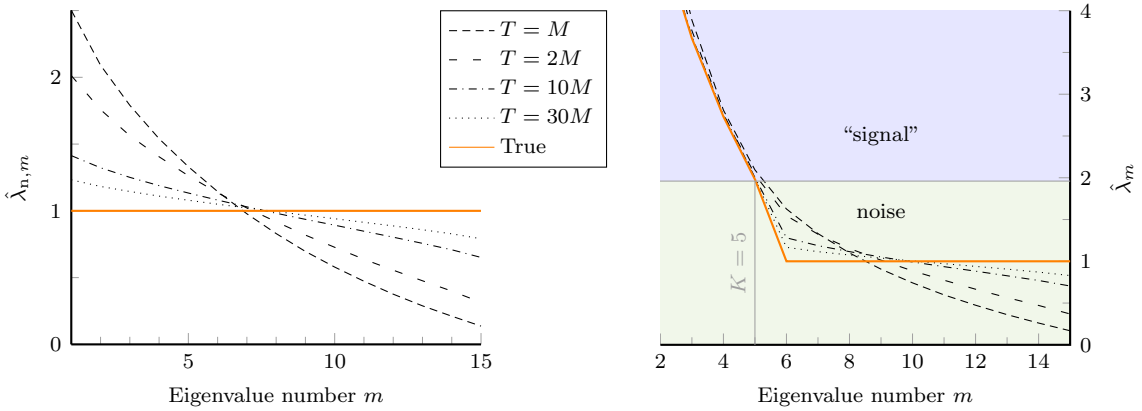
[2]Note that (4.2) implies that $T \to \infty$ so that the MMV model turns into a so-called IMV [137].

[3]In fact, the first K eigenvalues represent signal plus noise but the latter is routinely omitted for the sake of brevity. To avoid the confusion, we refer to the the first K eigenvalues of $\hat{\boldsymbol{R}}_{\mathrm{z}}$ as "signal" eigenvalues.

hypothesis testing as in [212] and [213], a predicted eigen-threshold approach [214], as well as the exponential fitting test (EFT) that exploits the exponentially decaying profile of the noise eigenvalues [215], [216]. A more in-depth discussion of different algorithms can be found in [204] and [205].

Example 4.12

To show how the limited number of snapshots affects the eigenvalue profile of the sample covariance matrix $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}}$, we generate an MMV setup from (4.1) with $N = 150$, $M = 15$ and $K = 5$ where $\boldsymbol{A} \sim \mathcal{CN}(0, 1)$, while the non-zero elements of $\boldsymbol{x}_t$ are drawn from $\mathcal{CN}\left(0, \sigma_{\mathrm{x}}^2\right)$. The noise $\boldsymbol{n}_t$ is standard complex normal, whereas the SNR is given by $\eta = K\sigma_{\mathrm{x}}^2$. Figure 4.1a demonstrates the ordered eigenvalues $\hat{\lambda}_{\mathrm{n},m}$ of the noise covariance matrix $\hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}}$ averaged over 10^2 realizations for several values of T together with their true (constant and equal to 1) value provided for reference. One can notice the typical exponential profile exhibited by the noise eigenvalues where the deviation from the true value increases with the decrease of the number of snapshots T. The respective eigenvalue profiles of $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}}$ are depicted in Figure 4.1b. The eigenvalues with the index number greater than $K = 5$ correspond to the noise subspace of $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}}$, whereas the ones with the lower index belong to its "signal" subspace. One can see that the distinction between the two areas becomes less pronounced at low values of T. One reason behind this is the aforementioned exponential decay of the noise eigenvalues. Another reason is the "leakage" of the signal energy from the first K "signal" eigenvalues to the last $(M - K)$ noise ones via the non-zero cross term $\hat{\boldsymbol{R}}_{\mathrm{x},\tilde{\mathrm{n}}}$. Both effects intensify with the decrease of T, which impedes SOE under small number of snapshots conditions.



(a) Noise eigenvalues $\hat{\lambda}_{\mathrm{n},m}$ **(b)** Total eigenvalues $\hat{\lambda}_m$

Figure 4.1: Influence of the number of snapshots on the eigenvalue profile for the case of a Gaussian $\boldsymbol{A}$ with $N = 150$ and $\rho = 0.1$ where $K = 5$ and $\eta = -17$ dB: the noise eigenvalue profile $\hat{\lambda}_{\mathrm{n},m}$ (a) and the total eigenvalue profile $\hat{\lambda}_m$ (b).

4.2.2.2 Empirical eigenvalue-threshold test (ETT)

Noise covariance matrix estimation. In order to perform prewhitening in (4.5), the noise covariance matrix $\boldsymbol{\Sigma}$ has to be known. For instance, if both the signal and the measurement noise are known to be white with elements that have known variances σ_{s}^2 and σ_{m}^2, it can be computed according to (3.1). However, practically the noise statistics is often unknown a prior and has to be estimated instead. Doing so requires collecting a training set $\boldsymbol{N}^{\mathrm{tr}} = [\boldsymbol{n}_1^{\mathrm{tr}}, \boldsymbol{n}_2^{\mathrm{tr}}, \ldots, \boldsymbol{n}_{T_{\mathrm{tr}}}^{\mathrm{tr}}] \in \mathbb{C}^{M \times T_{\mathrm{tr}}}$ of noise only samples and estimating $\boldsymbol{\Sigma}$ as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T_{\mathrm{tr}}} \boldsymbol{N}^{\mathrm{tr}} \left(\boldsymbol{N}^{\mathrm{tr}}\right)^{\mathrm{H}}. \tag{4.9}$$

The training set $\boldsymbol{N}^{\mathrm{tr}}$ can be obtained during a calibration step from a portion of data that is known to contain only noise, e.g., in radio transmitters one can evaluate the receiver noise by placing matching loads on the antenna outputs and recording the resulting response; one can also use generated input signals, such as single harmonics, in a controlled environment and estimate the noise characteristics by comparing the known signal with the recorded data. Alternatively, the training can be performed beforehand based on synthetically created noise samples, which requires a model for the noise, such as (3.1) with a certain distribution for $\boldsymbol{n}_t$ for instance, to be available. Regardless of the particular calibration method, availability of such training data provides an opportunity for sparsity order estimation as it allows to evaluate statistical characteristics of the noise eigenvalues along with the noise covariance matrix $\boldsymbol{\Sigma}$. Below we introduce an empirical threshold test (ETT) that takes use of this training data for SOE.

Hypothesis testing for ETT. We begin with formulating the sparsity order estimation as a set of binary hypothesis tests. For each test eigenvalue $\hat{\lambda}_m$ of the sample covariance matrix $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}}$ we can write the following hypotheses

$$\begin{cases} \mathcal{H}_{0,m} & : \quad \hat{\lambda}_m \in \mathcal{P}_{\mathrm{n}} \\ \mathcal{H}_{1,m} & : \quad \hat{\lambda}_m \in \mathcal{P}_{\mathrm{s}}, \end{cases} \tag{4.10}$$

where $\mathcal{P}_{\mathrm{n}}$ and $\mathcal{P}_{\mathrm{s}}$ are sets of noise only and signal plus noise eigenvalues, respectively. Taking into account that the test eigenvalues $\hat{\lambda}_m$ are sorted in a descending order, the sparsity order K is then simply given by the number of the largest eigenvalue that belongs to the "signal" subspace or, in other words,

$$\hat{K} = \max_{m:\{\hat{\lambda}_m \in \mathcal{P}_{\mathrm{s}}\}} (m). \tag{4.11}$$

Given (4.10), we can use a classical Neyman-Pearson rule (NPR) [217, Ch. 3.3] to differentiate between $\mathcal{H}_{0,m}$ and $\mathcal{H}_{1,m}$. Note that the NPR detector maximizes the probability of correct detection P_{d} for a fixed probability of false alarm P_{fa} resulting in a so-called constant false alarm rate (CFAR) performance. Denoting by α_m the desired probability of

false alarm at each sample eigenvalue, the decision rule for (4.10) can be formulated as

$$\hat{\lambda}_m \underset{\mathcal{H}_{0,m}}{\overset{\mathcal{H}_{1,m}}{\gtrless}} \zeta_m, \quad \text{where} \quad \zeta_m = \bar{F}_{\mathcal{H}_{0,m}}(\alpha_m). \tag{4.12}$$

Here, $\bar{F}_{\mathcal{H}_{0,m}}(\alpha_m)$ is a CCDF of $\hat{\lambda}_m|\mathcal{H}_{0,m}$. Note that under $\mathcal{H}_{0,m}$ ("noise only") the PDF of the test eigenvalue $\hat{\lambda}_m$ is (asymptotically) given by the PDF of the corresponding noise eigenvalue such that $f_{\mathcal{H}_{0,m}}(\hat{\lambda}_m) \overset{T\uparrow}{\approx} f(\hat{\lambda}_{\mathrm{n},m})$. The approximation is due to the presence of the cross-term $\hat{\boldsymbol{R}}_{\mathrm{x},\tilde{\mathrm{n}}}$ in (4.7), which allows "leakage" of signal energy from signal eigenvalues into the noise eigenvalues as discussed in Example 4.12. When the input signal is unknown this effect cannot be accounted for. Therefore, in practice one uses $f(\hat{\lambda}_{\mathrm{n},m})$ as a proxy for $f_{\mathcal{H}_{0,m}}(\hat{\lambda}_m)$ so that $\zeta_m = \bar{F}_{\hat{\lambda}_{\mathrm{n},m}}(\alpha_m)$, where $\bar{F}_{\lambda_{\mathrm{n},m}}$ represents the CCDF of $\hat{\lambda}_{\mathrm{n},m}$.

The decision rule (4.12) operates with M thresholds η_m obtained from M individual CCDFs, one for each sample eigenvalue $\hat{\lambda}_m$. This in turns results in M individual probabilities of false alarm $P_{\mathrm{fa},m}$ and missed detection $P_{\mathrm{m},m}$ for each $\hat{\lambda}_m$. Such an approach however, can potentially create ambiguities for the "signal" subspace estimation since false alarms and missed detections can occur at each of the sample eigenvalues. To illustrate this, suppose that an M-length binary vector $\boldsymbol{h}$ contains the decisions for M sample eigenvalues such that its mth element is equal to 1 in case the hypotheses $f_{\mathcal{H}_{1,m}}(\hat{\lambda}_m)$ has been accepted for $\hat{\lambda}_m$, or it is 0 otherwise. In addition, denote by K_{all} the number of elements of $\boldsymbol{h}$ that encompass all ones, meaning that $h_{K_{\mathrm{all}}} = 1$ while $\forall j > K_{\mathrm{all}}\ h_j = 0$, and by K_{first} the (largest) number of consecutive ones in $\boldsymbol{h}$ starting from $m = 1$, i.e., for all $i \in [1, K_{\mathrm{first}}]\ h_i = 1$. If $K_{\mathrm{first}} = K_{\mathrm{all}}$, the estimate of the sparsity order can be clearly obtained[4] as $\hat{K} = K_{\mathrm{all}}$ in accordance to (4.11). However, the application of individual thresholds can also result in a situation when first K_{all} elements contain a mixture of zeros and ones due to the combination of false alarms and missed detections at sample eigenvalues with different indices. In this case, $K_{\mathrm{all}} \neq K_{\mathrm{first}}$ and direct usage of (4.11) might result in a significant order over-estimation. Alternatively, on might also think of adopting a strategy from [214] where $\hat{K} = K_{\mathrm{first}}$. This, however can potentially lead to the opposite effect - order underestimation.

Instead of setting m different thresholds ζ_m based on m individual PDFs $f(\hat{\lambda}_{\mathrm{n},m})$, we use an aggregate PDF[5] $f(\hat{\lambda}_{\mathrm{n}})$ to compute a unified threshold η (cf. Example (4.13)). In other words, $f(\hat{\lambda}_{\mathrm{n}})$ is an "aggregate" distribution for an unordered set of noise eigenvalues. As a result, (4.12) transforms into

$$\hat{\lambda}_m \underset{\mathcal{H}_{0,m}}{\overset{\mathcal{H}_{1,m}}{\gtrless}} \zeta, \quad \text{where} \quad \zeta = \bar{F}_{\lambda_{\mathrm{n}}}(\alpha). \tag{4.13}$$

Here, $\bar{F}_{\lambda_{\mathrm{n}}}$ denotes the CCDF of $\hat{\lambda}_{\mathrm{n}}$, whereas α is the total desired false alarm rate. Obtaining a single threshold removes the ambiguity discussed above as it provides a single cut-off value

[4] This does not necessarily mean that $K_{\mathrm{all}} = K$ as both missed detections and false alarms can still occur.

[5] The absence of an eigenvalue index number m indicates that $\hat{\lambda}_{\mathrm{n}}$ represents any eigenvalue of $\hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}}$.

below which all sample eigenvalues are judged as belonging to the noise subspace[6]. An additional advantage of (4.13) is that for a low snapshot size T it provides larger threshold values for eigenvalues with higher index than those provided by individual PDFs. This can be helpful for low T since, as discussed in Example 4.12, the sample eigenvalues belonging to the noise subspace tend to be larger than what is predicted from their test statistics due to the energy leakage between signal and noise subspaces.

Example 4.13

In this example, we show individual PDFs $f(\hat{\lambda}_{\mathrm{n},m})$ and an aggregate PDF $f(\hat{\lambda}_{\mathrm{n}})$ for a scenario from Example 4.12 where $T = M$. Both are presented in Figure 4.2, where $f(\hat{\lambda}_{\mathrm{n},m})$ is depicted with respect to the left-hand y-axis, while $f(\hat{\lambda}_{\mathrm{n}})$ is shown with respect to the right-hand one. Since $f(\hat{\lambda}_{\mathrm{n}})$ accounts for all noise eigenvalues, a unified threshold η is likely to have a higher value than η_m. The difference between the two decreases with the increase of T as $f(\hat{\lambda}_{\mathrm{n},m})$ and $f(\hat{\lambda}_{\mathrm{n}})$ converge (asymptotically) to a delta function.

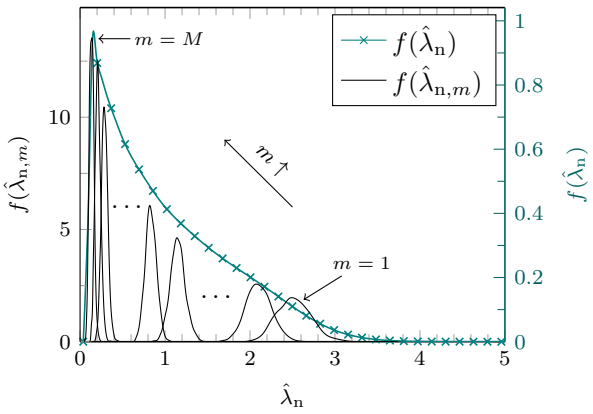


Figure 4.2: Individual (left-hand y-axis) and aggregate (right-hand y-axis) PDFs of noise eigenvalues.

Evaluation of noise eigenvalue PDF. Direct usage of (4.12) and (4.13) requires knowledge of the noise PDFs $f(\hat{\lambda}_{\mathrm{n},m})$ and $f(\hat{\lambda}_{\mathrm{n}})$, respectively. Although there is a large amount of results available for the asymptotic distributions of the sample noise eigenvalues $\hat{\lambda}_{\mathrm{n},m}$ for the case of white Gaussian noise [213], [214], [218] with some of these being also extended to the case of colored noise [218], the accuracy of such asymptotic results deteriorates when the signal dimensions are limited. Furthermore, the applicability of these results heavily depends on the validity of the assumptions about the noise statistics. The latter, on the other hand, can deviate from the assumed model, due to imperfect estimation of the noise covariance matrix $\boldsymbol{\Sigma}$ during the prewhitening step in (4.5) for instance. Using an empirical distribution of the noise eigenvalues obtained from the training noise samples $\boldsymbol{N}^{\mathrm{tr}}$ as an evaluation of $f(\hat{\lambda}_{\mathrm{n}})$ allows us to avoid this problem as it accounts for the actual noise statistics in a given setting (sensing matrix, number of snapshots, etc.). Below we show how $f(\hat{\lambda}_{\mathrm{n}})$ and the threshold η can be estimated from $\boldsymbol{N}^{\mathrm{tr}}$ step by step.

[6]Note that since $\tilde{\lambda}_m$ are ordered K_{all} is always equal to K_{first} in this case.

Given $\boldsymbol{N}^{\mathrm{tr}}$, we first calculate a training set of noise eigenvalue profiles $\hat{\boldsymbol{\lambda}}_{\mathrm{n}}^{(\ell)}$ as follows:

1. estimate the noise covariance matrix $\boldsymbol{\Sigma}$ according to (4.9);
2. compute whitened noise samples $\tilde{\boldsymbol{n}}_t^{\mathrm{tr}}$: $\tilde{\boldsymbol{n}}_t^{\mathrm{tr}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{n}_t^{\mathrm{tr}}$;
3. from $\tilde{\boldsymbol{n}}_t^{\mathrm{tr}}$ obtain $L_{\mathrm{tr}} = \lfloor \frac{T_{\mathrm{tr}}}{T} \rfloor$ sample noise covariance matrices $\hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}^{\mathrm{tr}}}^{(\ell)}$: $\hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}^{\mathrm{tr}}}^{(\ell)} = \frac{1}{T}\boldsymbol{N}_\ell^{\mathrm{tr}}\left(\boldsymbol{N}_\ell^{\mathrm{tr}}\right)^{\mathrm{H}}$, where $\boldsymbol{N}_\ell^{\mathrm{tr}} = [\tilde{\boldsymbol{n}}_{(\ell-1)T+1}^{\mathrm{tr}}, \ldots, \tilde{\boldsymbol{n}}_{\ell T}^{\mathrm{tr}}]$ and $\ell = 1, 2, \ldots, L_{\mathrm{tr}}$;
4. apply an eigenvalue decomposition to $\hat{\boldsymbol{R}}_{\tilde{\mathrm{n}}^{\mathrm{tr}}}^{(\ell)}$ to acquire L_{tr} noise eigenvalue profiles $\hat{\boldsymbol{\lambda}}_{\mathrm{n}}^{(\ell)}$.

After $\hat{\boldsymbol{\lambda}}_{\mathrm{n}}^{(\ell)}$ are obtained, we can form from them a single vector

$$\boldsymbol{\xi} = \left[\left(\hat{\boldsymbol{\lambda}}_{\mathrm{n}}^{(1)}\right)^{\mathrm{T}}, \ldots, \left(\hat{\boldsymbol{\lambda}}_{\mathrm{n}}^{(L_{\mathrm{tr}})}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{L\times 1},$$

where $L = ML_{\mathrm{tr}}$. From $\boldsymbol{\xi}$ we can now compute a histogram $P_q(\hat{\lambda}_{rmn})$ where $Q \in \mathbb{N}$ is the number of bins for the histogram estimation and $\tau = (\max_j(\xi_j) - \min_j(\xi_j))/Q$ denotes the bin width. The empirical (discrete) distribution $\hat{\lambda}_{\mathrm{n}}$ can then be estimated as

$$\hat{f}(\hat{\lambda}_{\mathrm{n}}) = \sum_{q=1}^{Q} \frac{P_q(\hat{\lambda}_{\mathrm{n}})}{\tau L}\delta\left(\xi - (q-0.5)\tau - \xi_{\min}\right), \tag{4.14}$$

where $\delta(\cdot)$ is a Dirac delta function, $\xi_{\min} = \min_j(\xi_j)$ and $P_q(\hat{\lambda}_{\mathrm{n}})$ is the number of elements of $\boldsymbol{\xi}$ that fall within the interval $[\xi_{\min} + \tau(q-1), \xi_{\min} + \tau q)$, while $q = 1, 2, \ldots, Q$.

Finally, the detection threshold for a decision rule in (4.13) can be found from the ordered set $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_L$ as $\xi_{\lfloor \alpha L \rfloor}$. Note that α asymptotically approaches the true probability of false alarm with increasing T and L_{tr}.

SOE performance evaluation. To evaluate the performance of ETT, we compare it against its counterpart MOS algorithms as a function of the SNR, the number of snapshots taken for covariance matrix estimation T, and the size of the training set L_{tr} in the setting of (4.7) . More specifically, we consider two common MOS algorithms: the non-parametric information theoretic approach EDC [211] and the statistical method EFT [214] that is parameterized by the desired false alarm rate $P_{\mathrm{fa}}^{\mathrm{t}}$. The parameters of (4.1) are set according to those of Examples 4.12 and 4.13; namely, Gaussian $\boldsymbol{A}$ with $N = 150$ and $\rho = 0.1$ (i.e., $M = 15$). The sparsity order K is chosen uniformly at random from $[1, 5]$ while the SNR is defined as $\eta = \frac{K\sigma_{\mathrm{x}}^2}{M\sigma_0^2}$ and $L_{\mathrm{tr}} = 100$. This default setting is summarized in Table 4.1.

Parameter	N	M	K	T	α	L_{tr}	ς	η	**Estimation algorithms**
Default value	150	15	$[1, 5]$	15	10^{-2}	100	1	0 dB	ETT, EFT, EDC

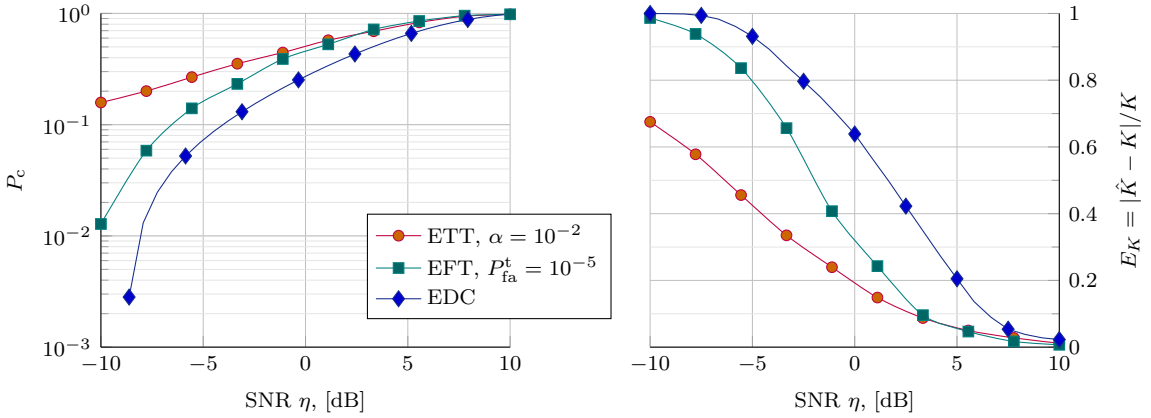
Table 4.1: List of parameters for SOE performance evaluation.

We begin with examining the SOE performance with respect to the SNR. Figure 4.3a shows the probability of correct estimation $P_{\mathrm{c}} = \Pr\left[\hat{K} = K\right]$ for three considered algorithms with $T = M$ and $\varsigma = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{m}}^2} = 1$. The default values of α and $P_{\mathrm{fa}}^{\mathrm{t}}$ for ETT and EFT are given by 10^{-2} and 10^{-5}, respectively. We see that both ETT and EFT outperform EDC, with ETT being slightly superior to EFT in the low SNR range. Looking at the mean estimation error $E_K = \frac{|\hat{K}-K|}{K}$ presented in Figure 4.3b we observe a similar pattern. To investigate the balance between the tendencies of the algorithms to over- and underestimate the sparsity order, Figure 4.4 shows the so-called receiver operating characteristics (ROC). Figure 4.4a depicts empirical probability of missed detection $P_{\mathrm{m}} = \Pr\left[\hat{K} < K\right]$ vs. empirical probability of false alarm $P_{\mathrm{fa}} = \Pr\left[\hat{K} > K\right]$ for two SNR values, whereas Figure 4.4b demonstrates corresponding curves for P_{c} and P_{fa}. First, we note that EDC in both cases provides a single point on the ROC plane as it is a non-parametric method, and as such it has a fixed performance for any given SNR. Furthermore, its operating points are located in low false alarm/high missed detection area, which means that it consistently underestimates the sparsity order. Although this tendency for model order underestimation is fundamental to eigenvalue-based MOS algorithms [219], ETT and EFT allow a trade-off between the degrees of over- and underestimation by tuning the target false alarm rate. This is, however, generally not a feature of information theoretic approaches, which makes them less suited[7] for SOE. Regarding ETT and EFT, we can see that while they perform almost identically in the high SNR range, in the lower SNRs ETT provides higher estimation accuracy both in terms of underestimation rate and probability of correct order estimation.

In the next set of numerical simulations, we study the influence of the number of snapshots T taken for the estimation of the sample covariance matrix in (4.7) and the size of the training set T_{tr} used for the noise covariance matrix estimation in (4.9). Thus, Figures 4.5a and 4.5b show the probability of correct estimation P_{c} for an SNR of 0 dB as a function of T where $L_{\mathrm{tr}} = 100$ and as a function of $L_{\mathrm{tr}} = \frac{T_{\mathrm{tr}}}{T}$ where $T = M$, respectively. From the former we see that the performance of ETT and EDC predictably improves with the growth of the number of snapshots T, whereas for EFT it does so only for a short while, after which it starts to deteriorate. This is because EFT is designed for order estimation with a low number of snapshots where the noise eigenvalues exhibit pronounced exponential profile. Turning our attention to Figure 4.5b, we observe that the performance of ETT improves with larger training set, however rather mildly. This means that a relatively small number of training samples suffices to obtain reliable estimates of the noise covariance matrix and the decision threshold η in (4.13). Interestingly, the performance of EFT also slightly increases with larger training size, which can be explained by better noise equalization[8] during (4.5).

[7] In CS context, underestimation of the sparsity order is more critical that overestimation, especially when the estimated order is used to adapt the number of measurements or assist the recovery.

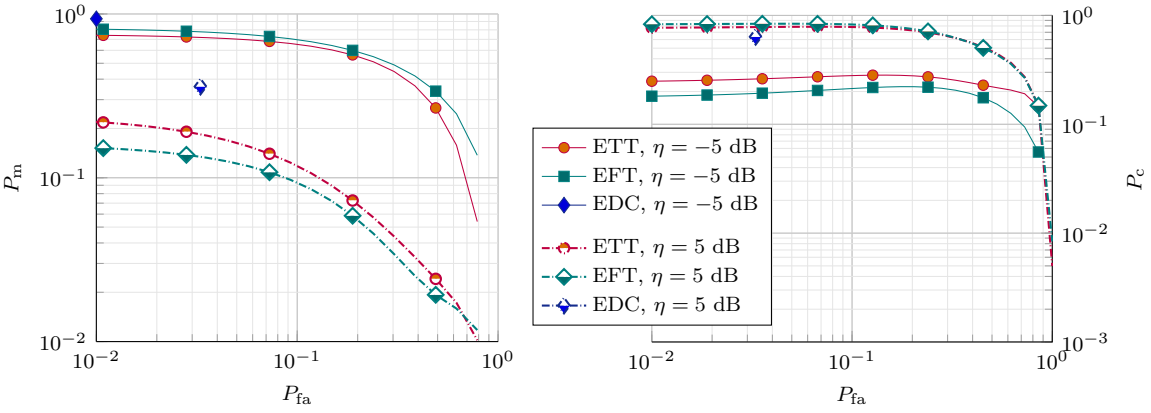
[8] The EFT assumes white noise model for its training step, which makes it sensitive to noise model mismatch.



(b) Probability of correct estimation

(a) Mean estimation error

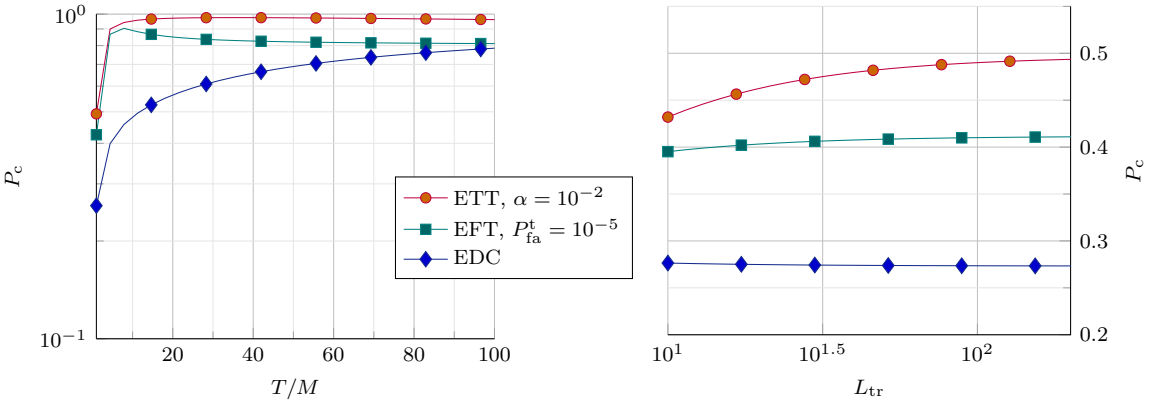
Figure 4.3: SOE performance with respect to SNR: empirical probability of correct estimation P_c (a) and mean sparsity order estimation error $E_K = \frac{\hat{K}-K}{K}$ (b).



(a) Missed detection vs. false alarm

(b) Correct detection vs. false alarm

Figure 4.4: Empirical ROC curves: probability of missed detection P_m vs. probability of false alarm P_{fa} (a) and probability of correct detection P_c vs. probability of false alarm P_{fa} (b).



(a) P_c vs. T/M

(b) P_c vs. L_{tr}

Figure 4.5: Probability of correct estimation P_c with respect to the number of snapshots T (a) and the size of the training set L_{tr} (b).

Presented results allow us to conclude that the two statistical methods, the proposed empirical threshold test (ETT) and the exponential fitting test (EFT) designed for MOS, provide comparable sparsity order estimation performance for low snapshot size T. They both allow a trade-off between over- and underestimation rates, which is a significant advantage for the considered application. The information-theoretic efficient detection criterion (EDC) on the other hand, proves to be less promising in the CS context as it consistently underestimates the order and is significantly more susceptible to noise.

Recovery performance evaluation. Finally, we investigate how the SOE impacts the sparse recovery. To do so, in Figure 4.6 we plot the probability of correct order estimation (P_{c}) vs. the probability of correct support recovery ($P_{\mathrm{sr}} = \Pr\left[\mathcal{S}(\hat{\boldsymbol{X}}) = \mathcal{S}(\boldsymbol{X})\right]$) where each point corresponds to a certain SNR value (smaller values of P_{c} and P_{sr} correspond to lower SNRs as indicated in the figure) while the model parameters are given in Table 4.1. For the support recovery step we use the simultaneous orthogonal matching pursuit (SOMP) algorithm [134] (see Algorithm 2 in Section 2.3.2) where as a stopping criterion we use the ratio, denoted by ε, of the residual norm to the norm of the current solution. For comparison, we show the results for several different values of ε and highlight by red the area of optimal support recovery performance[9] that corresponds to $\varepsilon \in [0.2, 0.3]$. Note that it is optimal in the sense that in the high SNRs the support recovery rate reaches 1 while in the lower SNRs it is minimized. We can see that the latter is a relatively narrow strip located largely in the upper left corner of the graph. This indicates that reliable order estimation is achieved before (for lower SNRs) than reliable support recovery and hence it can be used for recovery assistance. As an example, we also plot P_{sr} for the case when the stopping criterion is the true sparsity order[10] K. In contrast, it lies below the the upper triangle closer to the right-hand

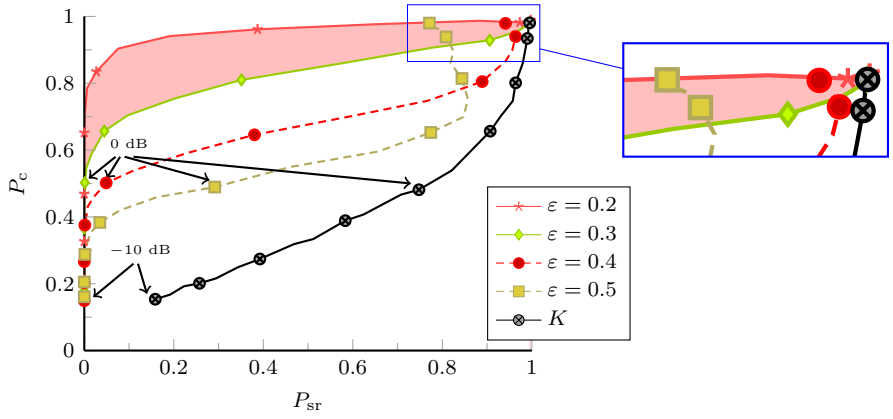


Figure 4.6: Probability of correct sparsity order estimation vs. probability of correct support recovery.

[9] Finding an optimal value for a stopping criterion based on the residual is a non-trivial task that is often solved by a "trial and error" approach. An optimal strategy would be to adapt ε according to the current SNR, which however is often impractical since it requires knowledge of the SNR that is seldom available.

[10] Under certain conditions, OMP is guaranteed to recover any K-sparse vector in exactly K iterations [57].

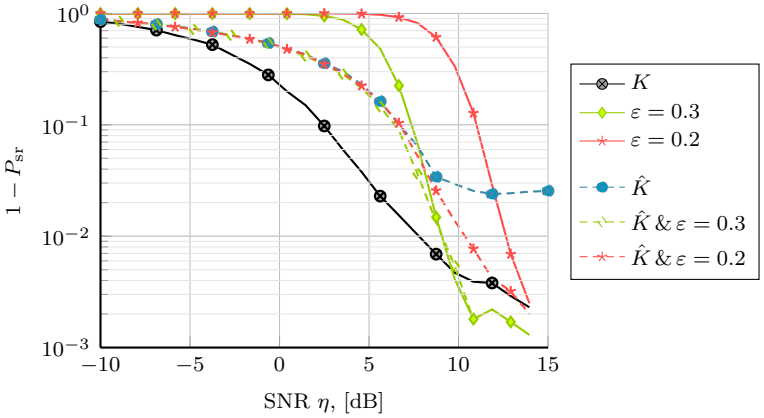


Figure 4.7: Probability of erroneous support recovery vs. SNR with or without SOE.

corner, which shows the performance improvement that can be achieved by knowing the exact value of K. The corresponding curves with respect to SNR are provided in Figure (4.7). Here we plot $(1 - P_{\text{sr}})$ vs. SNR for three cases considered before, namely, when the stopping criterion is the true sparsity order K or the norm ratio $\varepsilon = 0.2, 0.3$. Additionally, we show the results for the case when the stopping criterion is the sparsity order estimate $\hat{K}$ obtained via ETT with $\alpha = 10^{-2}$, either alone (marked by $\hat{K}$) or in combination with the norm ratio ε (marked by $\hat{K}\,\&\,\varepsilon = k$). We see that using $\hat{K}$ as a stopping criterion significantly improves the support recovery performance, especially in the low SNRs. Due to the non-zero false alarm probability it alone does not however allow for perfect recovery in the high SNRs. Combining both criteria (sparsity order estimate with the residual one), we can eliminate this problem and enjoy the improved performance in the low SNRs provided by SOE.

4.3 Detection of time-varying support

4.3.1 Quasi-stationary MMV

In previous section, we discussed the sparsity order estimation in an MMV setting where a number of consecutive input signals form a jointly K-sparse input matrix $\boldsymbol{X}$ which is later compressively acquired to obtain a matrix of measurements $\boldsymbol{Y}$. Such a model finds its place in a variety of practical applications. However, it is often the case that the support of the input signal exhibits variations from one block of data to another, preserving the jointly sparse structure of (2.31) for a limited number of snapshots only, as illustrated in Figure 4.8. Then, we can write the (noise-free) matrix of measurements $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_L] = \boldsymbol{A}\boldsymbol{X} = [\boldsymbol{A}\boldsymbol{X}_1, \boldsymbol{A}\boldsymbol{X}_2, \ldots, \boldsymbol{A}\boldsymbol{X}_L], \tag{4.15}$$

where $\boldsymbol{Y}_k = \boldsymbol{A}\boldsymbol{X}_k \in \mathbb{C}^{M \times t_k}$ and $|\mathcal{S}(\boldsymbol{X}_k)| \leq K$, while $k = 1, \ldots, L$ and $L : \sum_{k=1}^{L} t_k = T$. Note that the individual supports $\mathcal{S}(\boldsymbol{X}_k)$ are not equal meaning that $\mathcal{S}(\boldsymbol{X}_k) \neq \mathcal{S}(\boldsymbol{X}_{k+1}) \forall k =$

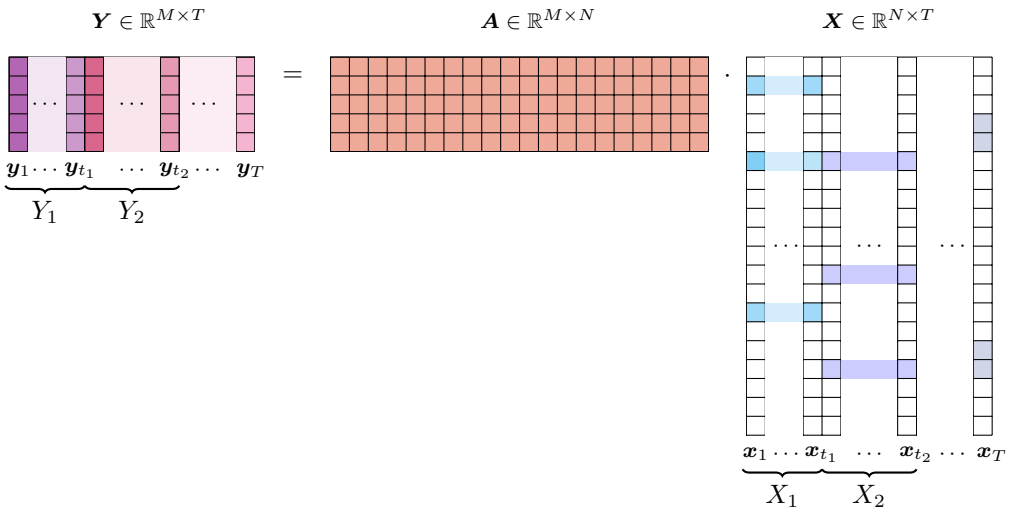


Figure 4.8: Finite-dimensional quasi-stationary MMV system with an $M \times T$ matrix $\boldsymbol{Y}$ containing measurements of T K-sparse signals $\boldsymbol{x}_t$ that are acquired via the application of the sensing matrix $\boldsymbol{A}$. Consecutive signals $\boldsymbol{x}_t$ form jointly K-sparse matrices X_k of length t_k such that $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_L]$.

$[1, L-1]$. We call the MMV model with such a varying support *quasi-stationary*.

Once could solve (4.15) as a regular MMV problem, e.g., via

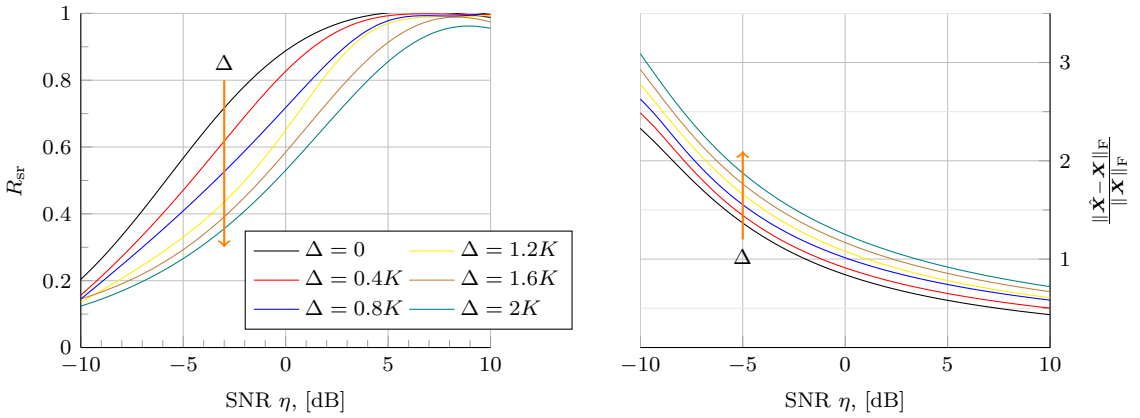
$$\hat{\boldsymbol{X}} = \arg\left(\min_{\boldsymbol{X}} |\mathcal{S}(\boldsymbol{X})|\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}. \tag{4.16}$$

However, we note that due to the composite quasi-stationary structure of $\boldsymbol{X}$ the cardinality of $\mathcal{S}(\boldsymbol{X})$, given by a union over the supports of sub-matrices $\boldsymbol{X}_k$

$$\mathcal{S}(\boldsymbol{X}) = \bigcup_{k=1}^{L} \mathcal{S}(\boldsymbol{X}_k), \tag{4.17}$$

is greater than that of each $\boldsymbol{X}_k$ individually. More specifically, it can reach up to $\min(LK, N)$ in case of all blocks having non-overlapping supports, i.e., $K \leq |\mathcal{S}(\boldsymbol{X})| \leq \min(LK, N)$. This means that the effective sparsity order of $\boldsymbol{X}$ increases, diminishing the advantages offered by the joint sparse structure (see Example 4.14). In the worst case scenario, the sparsity order of $\boldsymbol{X}$ can exceed the recoverability conditions for a given number of measurements M, which will result in the complete signal recovery collapse. Contrarily, if the block lengths t_k, which we refer to as the stationarity windows of $\boldsymbol{X}$, are perfectly known, the recovery of $\boldsymbol{X}$ can be formulated as a set of regular MMV problems, i.e.,

$$\hat{\boldsymbol{X}} : \begin{cases} \hat{\boldsymbol{X}}_1 = \arg\left(\min_{\boldsymbol{X}_1} |\mathcal{S}(\boldsymbol{X}_1)|\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{X}_1 = \boldsymbol{Y}_1 \\ \quad \vdots \\ \hat{\boldsymbol{X}}_L = \arg\left(\min_{\boldsymbol{X}_L} |\mathcal{S}(\boldsymbol{X}_L)|\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{X}_L = \boldsymbol{Y}_L \end{cases}, \tag{4.18}$$



(a) Support recovery rate

(a) Root mean squared error

Figure 4.9: Support recovery rate R_{sr} (a) and RMSE (b) for direct MMV recovery according to (4.16) with an undetected support change. The value of $\Delta = 0$ corresponds to the recovery of the static MMV without the support change.

where $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{X}}_1, \ldots, \hat{\boldsymbol{X}}_L]$, whereas $\boldsymbol{Y}_1 = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t_1}]$ and $\boldsymbol{Y}_L = [\boldsymbol{y}_{\sum_{k=1}^{L-1} t_k+1}, \ldots, \boldsymbol{y}_T]$. Application of (4.18) makes use of MMV structure of individual blocks $\boldsymbol{X}_k$ and avoids the pitfalls of potential sparsity order increase due to the changing sparsity pattern. However, it requires the knowledge of the stationarity intervals t_k.

Example 4.14

In this example, we demonstrate performance degradation caused by the undetected sparsity pattern change in MMV. We simulate a quasi-stationary MMV scenario with $N = 150$, $M = 15$ and $K = 5$ where $\boldsymbol{X}$ contains $L = 2$ segments of length $t_1 = t_2 = M$. We then vary the size $\Delta = |\mathcal{S}(\boldsymbol{X})| - |\mathcal{S}(\boldsymbol{X}_1) \bigcap \mathcal{S}(\boldsymbol{X}_2)|$ of the support difference between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ and compute the support recovery rate $R_{\text{sr}} = \frac{|\mathcal{S}(\boldsymbol{X}) \cap \mathcal{S}(\hat{\boldsymbol{X}})|}{|\mathcal{S}(\boldsymbol{X})|}$ and the normalized RMSE $\frac{\|\hat{\boldsymbol{X}}-\boldsymbol{X}\|_{\text{F}}}{\|\boldsymbol{X}\|_{\text{F}}}$ using the SOMP algorithm for MMV recovery according to (4.16).

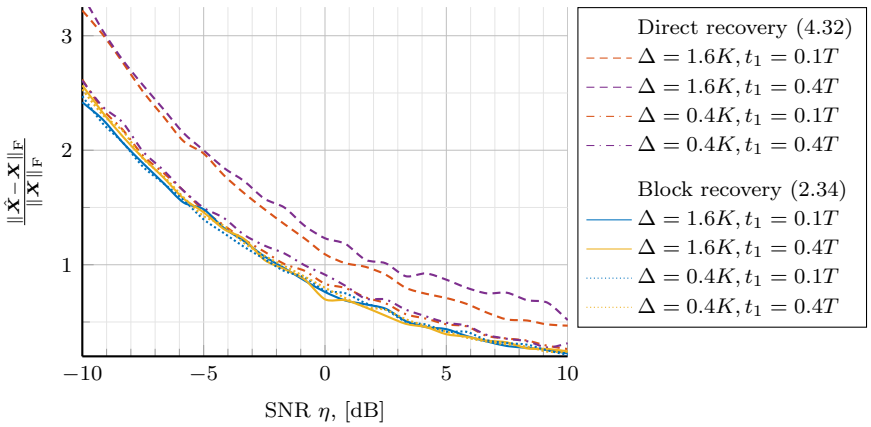


Figure 4.10: Comparison of recovery strategies for quasi-stationary MMV.

Figure 4.9a and b show resulting R_{sr} and RMSE as functions of SNR for different values of Δ, respectively. We observe that the recovery performance in therms of both metrics degrades with the increase of the difference between the supports. In the worst case, when $\Delta = 2K$ and the supports of two segments do not overlap at all, the performance degradation compared to a static MMV with $\Delta = 0$ is equivalent to ≈ 5 dB. Additionally, the support recovery rate in this case reaches a plateau at a level below 100% recovery.

Next, in Figure 4.10 we compare two recovery strategies, the direct MMV recovery according to (4.16) and the MMV recovery of individual blocks according to (4.18). We observe that the knowledge of the support change and subsequent recovery of the static MMV blocks with coinciding support provides consistently better estimation accuracy compared to the direct recovering of an entire quasi-static MMV at once. Obviously, the improvement grows with Δ and t_1 as the performance of direct recovery deteriorates.

In previous section, we have shown that in the static MMV setup the signal sparsity order can be assessed via the estimation of the effective rank of the block of measurement data, provided that the individual signal vectors are incoherent to each other. Capitalizing on these results, we propose to use the rank information in order to estimate the moment of the support change in the quasi-static scenario as well. To do so, we analyze the relation between the resulting ranks of the consequent windows of measurement data and the supports of the corresponding blocks of input data depending on the size and the position of the blocks.

4.3.2 Rank evolution analysis

4.3.2.1 General idea

Consider a scenario where the support changes only once[11]. The input matrix $\boldsymbol{X}$ consists then of two jointly K-sparse sub-matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of sizes $N \times t_1$ and $N \times t_2$, respectively, where

$$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2] \in \mathbb{C}^{N \times T} \text{ while } |\mathcal{S}(\boldsymbol{X}_1)| = |\mathcal{S}(\boldsymbol{X}_2)| = K : \ \mathcal{S}_{\text{X}_1} \neq \mathcal{S}_{\text{X}_2}, \tag{4.19}$$

with $T = t_1 + t_2$. As before, we assume a full-rank condition on both $\boldsymbol{X}$ and $\boldsymbol{X}_k$, which means that $\operatorname{rank}(\boldsymbol{X}) = |\mathcal{S}(\boldsymbol{X})| > \operatorname{rank}(\mathcal{S}(\boldsymbol{X}_k)) = |\mathcal{S}(\boldsymbol{X}_k)|$. Additionally, we assume that the individual vectors $\boldsymbol{x}_i$ within each block $\boldsymbol{X}_k$ have exactly[12] the same support meaning that $\mathcal{S}(\boldsymbol{x}_i) = \mathcal{S}(\boldsymbol{x}_j)$ for any $i, j \in [1 + \sum_{k=1}^{n-1} t_k, \sum_{k=1}^{n} t_k]$ where $n = 1, 2$ and $t_0 = 0$.

Now we define a windowed block of measurement data of length $2 \leq m \leq T$ as

$$\boldsymbol{Y}_p^{\text{w}} = \boldsymbol{A}\boldsymbol{X}_p^{\text{w}}, \tag{4.20}$$

[11] As an extension to an arbitrary number of support changes is rather straightforward we do not consider it here for the sake of brevity.

[12] Note that according to (2.30), a classical MMV does not necessarily require that all individual input vectors have the same support, but rather that the cardinality of the total support of $\boldsymbol{X}$ does not exceed K. This means that some of the input vectors can have a smaller support that is a subset of the total support set $\mathcal{S}(\boldsymbol{X})$. We discuss what happens in this case at the end of this section.

where $\boldsymbol{Y}_p^{\mathrm{w}} = [\boldsymbol{y}_p, \boldsymbol{y}_{p+1}, \ldots, \boldsymbol{y}_{p+m-1}]$ and $\boldsymbol{X}_p^{\mathrm{w}} = [\boldsymbol{x}_p, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_{p+m-1}]$, $p \in [1, T-m+1]$. The window size $m > 1$ is a parameter we are free to choose. The key idea here is that under certain conditions on the window size m the support of $\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})$ and, more importantly, its cardinality $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ will experience a change(s) caused by the transition from $\boldsymbol{X}_1$ to $\boldsymbol{X}_2$. The moment of this change(s) (the sliding index p at which the change happens) will in turn contain information on the value of t_1. Therefore, if we were able to observe this evolution of the support change through the evolution of the rank of the corresponding block of measurement data $\boldsymbol{Y}_p^{\mathrm{w}}$, we could identify t_1.

Consider the rank of $\boldsymbol{Y}_p^{\mathrm{w}}$. Clearly, we can bound it as follows

$$\operatorname{rank}\left(\boldsymbol{Y}_p^{\mathrm{w}}\right) \leq \operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right) \leq |\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|. \tag{4.21}$$

Note that, in contrast to the static MMV considered before, the rank of a windowed portion of the input data is not necessarily equal to its support size or the rank of the corresponding portion of the measurement data. Therefore, in order to use the sparsity order estimation approach for sparsity pattern change detection we first need to determine: 1) the conditions on the sensing matrix $\boldsymbol{A}$ and the window size m under which a change in the support of $\boldsymbol{X}_p^{\mathrm{w}}$ caused by the transition from $\boldsymbol{X}_1$ to $\boldsymbol{X}_2$ uniquely corresponds to a change in the rank of $\boldsymbol{Y}_p^{\mathrm{w}}$; 2) exact relations between the moments of these changes.

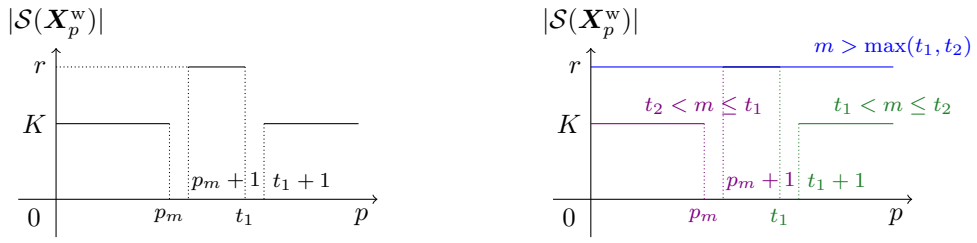
Consider first the left-hand part of (4.21). From Theorem 4, it turns into an equality when the rank of $\boldsymbol{A}$ is at least $\max_p \left(\operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right)\right)$. On the other hand, the maximum rank that $\boldsymbol{X}_p^{\mathrm{w}}$ can reach is $\min(K_\cup, m)$ where $K_\cup = |\mathcal{S}(\boldsymbol{X}_1) \cup \mathcal{S}(\boldsymbol{X}_2)| \leq 2K$. Hence, a sufficient condition for $\operatorname{rank}\left(\boldsymbol{Y}_p^{\mathrm{w}}\right) = \operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ is given by

$$\operatorname{rank}\left(\boldsymbol{A}\right) > \min(K_\cup, m). \tag{4.22}$$

Compared to (2.34), (4.22) introduces an additional term of $\min(K_\cup, m) - K$ to the rank requirements of $\boldsymbol{A}$, provided that $m \geq K$. Although, in the noise-free case this would mean an increase in the minimal required number of measurements, in practice this requirement is likely to be fulfilled as one normally would choose a larger number of measurements than required by the minimal conditions in order to account for the noise. Therefore, henceforth we suppose that (4.22) holds and $\operatorname{rank}\left(\boldsymbol{Y}_p^{\mathrm{w}}\right) = \operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ for any $p \in [1, T-m+1]$ and $m \in [2, T]$. The right-hand part of (4.21) we analyze in two steps: we study the value of $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ first and then establish its relationship to the rank of $\boldsymbol{X}_p^{\mathrm{w}}$.

4.3.2.2 Window support evolution

Consider the support of $\boldsymbol{X}_p^{\mathrm{w}}$. If $\boldsymbol{X}_p^{\mathrm{w}}$ contains only input vectors belonging to $\boldsymbol{X}_1$, its support is equal to the support of $\boldsymbol{X}_1$. Similarly, if it is comprised of the input vectors with index larger than t_1, its support is equal to that of $\boldsymbol{X}_2$. The former happens for any $m < t_1$ and $p \leq p_m = t_1 - m + 1$, whereas for the latter we need m to be smaller than t_2 and p greater



(a) $2 \leq m \leq \min(t_1, t_2)$ **(b)** $m > \min(t_1, t_2)$

Figure 4.11: Cardinality of the support set $\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})$ versus sliding index p where $p_m = t_1 - m + 1$ for $2 \leq m \leq \min(t_1, t_2)$ (a) and $m > \min(t_1, t_2)$ (b).

than t_1. On the other hand, when $p_m < p < t_1$ and $m \geq 2$, $\boldsymbol{X}_p^{\mathrm{w}}$ contains input vectors from both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, which means that its support is a union between $\mathcal{S}(\boldsymbol{X}_1)$ and $\mathcal{S}(\boldsymbol{X}_2)$. It is easy to see that the same is true for the cases when $p < t_1$ with $t_1 < m < t_2$ and $p > p_m$ with $t_2 < m < t_1$. Combining these together we can formulate a following proposition.

Proposition 2. *Let* $\boldsymbol{X}_p^{\mathrm{w}} = [\boldsymbol{x}_p, \dots \boldsymbol{x}_{p+m-1}]$ *be an* $N \times m$ *sub-matrix of an* $N \times T$ *matrix* $\boldsymbol{X} = [\boldsymbol{x}_1, \dots \boldsymbol{x}_T]$ *defined according to* (4.19) *where* $\boldsymbol{x}_k$ *is a* k*-th column of* $\boldsymbol{X}$ *and denote by* $\beta_i = \begin{cases} K, & \text{if } m \leq t_i \\ K_\cup, & \text{otherwise} \end{cases}$ *where* $i = 1, 2$*. Then the cardinality of* $\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})$ *is given by*

$$|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = \begin{cases} \beta_1, & p \in [1, p_m] \\ K_\cup, & p \in [p_m + 1, t_1], \\ \beta_2, & p \in [t_1 + 1, T - m + 1] \end{cases} \tag{4.23}$$

Using Proposition 2, we can identify four distinct cases depending on the value of m:

1. $2 \leq m \leq \min(t_1, t_2)$ (shown in Figure 4.11a): $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = \begin{cases} K, & \text{if } \begin{cases} p \leq p_m \\ p > t_1 \end{cases} \\ K_\cup, & \text{if } p_m < p \leq t_1 \end{cases}$.
2. $t_2 < m \leq t_1$ (shown in Figure 4.11b in violet color): $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = \begin{cases} K, & \text{if } p \leq p_m \\ K_\cup, & \text{if } p_m < p \end{cases}$.
3. $t_1 < m \leq t_2$ (shown in Figure 4.11b in green color): $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = \begin{cases} K_\cup, & \text{if } p \leq t_1 \\ K, & \text{if } p > t_1 \end{cases}$.
4. $m > \max(t_1, t_2)$ (shown in Figure 4.11b in blue color): $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = K_\cup$ for any $p \in [1, T - m + 1]$.

In the first three cases, $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ is a piece-wise constant function of p with at least one change of value whose index p_{ch} is defined by the stationarity window t_1; when the change is positive and $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ increases $p_{\mathrm{ch}} = p_m = t_1 - m + 1$, otherwise $p_{\mathrm{ch}} = t_1$. On the contrary, when $m > \max(t_1, t_2)$ the value of $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ exhibits no changes and $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = K_\cup$ irrespective of

p. Clearly, this case has no use for the purpose of support change detection, which leaves us with the interval $[2, \max(t_1, t_2)]$ of potential values of m.

Having analyzed the behavior of $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ with respect to p and m, we now turn to determining how it translates into the behaviour of the rank of $\boldsymbol{X}_p^{\mathrm{w}}$ and, subsequently, into the rank of $\boldsymbol{Y}_p^{\mathrm{w}}$.

4.3.2.3 Window rank evolution

Since $\boldsymbol{X}$ and any of its sub-matrices are full rank, we can write that

$$\operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right) = \begin{cases} \bar{\beta}_1, & p \in [1, p_m] \\ \alpha, & p \in [p_m + 1, t_1], \\ \bar{\beta}_2, & p \in [t_1 + 1, T - m + 1] \end{cases} \quad \text{where } \bar{\beta}_i = \begin{cases} \min(K, m), & \text{if } m \leq t_i \\ \alpha, & \text{otherwise} \end{cases} \tag{4.24}$$

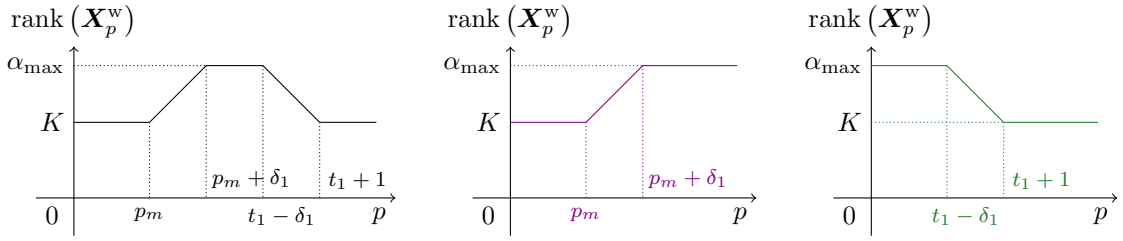
with $i = 1, 2$ and $\alpha \leq |\mathcal{S}(\boldsymbol{X}_1) \bigcup \mathcal{S}(\boldsymbol{X}_2)| \in \mathbb{N}$. We notice that (4.24) and (4.23) look similar except for the change from r to α and from β_i to $\bar{\beta}_i$. The latter become equal when $m \geq K$ and $m \leq t_i$. As for α, we investigate its value in the following.

To begin, we introduce some additional notations useful in the ensuing discussion:

- the redundancy $d_{\mathrm{mk}} = \begin{cases} m - K, & \text{if } m > K \\ 0, & \text{otherwise} \end{cases}$ of the window size m with respect to K.
- the size d_{21} of the difference set $\mathcal{S}(\boldsymbol{X}_1) \backslash \mathcal{S}(\boldsymbol{X}_2)$ where $A \backslash B$ denotes the relative compliment of B in A.
- the Heaviside step function $H[n] = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$.

Now we are ready to proceed with the analysis. Note that, although from the support change perspective m can be as small as 2, for the rank estimation a window size smaller than K is of no use since it does not preserve the sparsity order in the window rank. Furthermore, we can also immediately rule out $m = K$ as it would result in $\operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right) = K$ for any sliding index p. Therefore, henceforth we consider $m > K$ only.

Window size $m \in (K, \min(t_1, t_2)]$. When $m > K$ we have that $\operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right) = |\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})| = K$ for $p \leq p_m$ and $p \geq t_1 + 1$. The values of $\operatorname{rank}\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ in the interval $p_m < p \leq t_1$ can however differ significantly from those of $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$. To see why, consider the rank of $\boldsymbol{X}_{p_m+1}^{\mathrm{w}}$ for instance. Adding one vector from $\boldsymbol{X}_2$ to $m - 1$ vectors of $\boldsymbol{X}_1$ can increase the rank of $\boldsymbol{X}_{p_m+1}^{\mathrm{w}}$ by one only, irrespective of the difference between the supports. The rank hence will increase by one with each increment of p until it reaches $K_{\cup}$ at $p = p_m + d_{21}$ if $d_{\mathrm{mk}} \geq d_{21}$. It will stay equal to $K_{\cup}$ until $p = t_1 - d_{21}$, after which it will start to decrease till it reaches K



(a) $K < m \leq \min(t_1, t_2)$ **(b)** $t_2 < m \leq t_1$ **(c)** $t_1 < m \leq t_2$

Figure 4.12: Rank of $\boldsymbol{X}_p^{\mathrm{w}}$ versus sliding index p for $K < m \leq \min(t_1, t_2)$ (a), $t_2 < m < t_1$ (b) and $t_1 < m < t_2$ (c) where $p_m = t_1 - m + 1$, $\delta_1 = \min(d_{21}, d_{\mathrm{mk}})$ and $\alpha_{\max} = K_{\cup} + (d_{\mathrm{mk}} - d_{21})H[d_{21} - d_{\mathrm{mk}}]$.

at $p = t_1 + 1$, as demonstrated in Figure 4.12a. If $d_{\mathrm{mk}} < d_{21}$, it will increase until it reaches $K_{\cup} - d_{21} + d_{\mathrm{mk}}$ at $p = p_m + d_{\mathrm{mk}}$, after which it will start immediately decreasing.

Window size $m \in (t_2, t_1]$ or $m \in (t_1, t_2]$. The value of rank $\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ for $m \in (t_2, t_1]$ will stay equal to K up to $p = p_m$ as before and will start increasing from $p = p_m + 1$ until p reaches $p_m + \delta_1$ (see Figure 4.12b). Note that this corresponds to the first part of the curve for rank $\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ with $m \in (K, \min(t_1, t_2)]$. Similarly, when $m \in (t_1, t_2]$ the value of rank $\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ behaves according to the second half of Figure 4.12a (see Figure 4.12c).

Based on the presented analysis, we can formulate a following proposition.

Proposition 3. *Let $\boldsymbol{X}_p^{\mathrm{w}} = [\boldsymbol{x}_p, \ldots \boldsymbol{x}_{p+m+1}]$ be an $N \times m$ sub-matrix of an $N \times T$ matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_T]$ defined according to* (4.19) *where $\boldsymbol{x}_k$ is the k-th column of $\boldsymbol{X}$. Then, if $K < m \leq \max(t_1, t_2)$ there exist at least one index p such that* rank $\left(\boldsymbol{X}_p^{\mathrm{w}}\right) \neq$ rank $\left(\boldsymbol{X}_{p+1}^{\mathrm{w}}\right)$ *and $p = t_1 + \delta$ where δ does not depend on any other system parameters but (possibly) m.*

Proposition 3 provides a condition on the window size m that ensures the presence of a change in the rank of a sliding window $\boldsymbol{X}_p^{\mathrm{w}}$ whose moment is determined by the unknown value of t_1 and the known value of m. Furthermore, the behavior exhibited by rank $\left(\boldsymbol{X}_p^{\mathrm{w}}\right)$ essentially repeats that of $|\mathcal{S}(\boldsymbol{X}_p^{\mathrm{w}})|$ except for the presence of monotonic slopes between the constant-value regions, as demonstrated in Figure 4.12. The number, length and type of the slopes are determined by the relation between m, t_1 and t_2. In contrast to the case of the support evolution, not all the moments of change in the rank of $\boldsymbol{X}_p^{\mathrm{w}}$ are solely determined by t_1 and m, some also depend on the relation between the supports of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

4.3.3 Stationarity interval estimation

4.3.3.1 Proposed approach

Given Proposition 3 and Figure 4.12, we could formulate the following rule for the estimation of t_1 in the noise-free case. When (4.22) holds and $K < m \leq \max(t_1, t_2)$, the value of t_1 can be found by identifying the index p that corresponds to the first and/or last change in the

values of the rank of $r_p = \text{rank}\left(\boldsymbol{Y}_p^{\text{w}}\right)$ according to

$$t_1 = \begin{cases} p_{\text{fc}} + m - 1 & \text{, if } \forall p \in [p_{\text{fc}}, p_{\text{lc}}] : r_p < r_{p+1} \\ p_{\text{lc}} - 1 & \text{, otherwise} \end{cases}, \tag{4.25}$$

where p_{fc} and p_{lc} denote the indices of the first and last change in r_p, respectively.

A direct application of (4.25) for stationarity interval estimation (SIE) is nevertheless complicated by several practical considerations. First, the window rank (i.e., the sparsity order) r_p has to be estimated from a set of noisy measurements which implies the possibility of erroneous estimates. Second, we need to ensure that the window size m is within the proper range, i.e., $m \in (K, \max(t_1, t_2)]$. One possible solution would be to set m close to its lower bound K. We have already seen however that in the presence of noise, low number of snapshots creates difficulties for reliable rank estimation, increasing the probability of obtaining erroneous estimates. Hence, from the perspective of reliable rank estimation one would like to choose m close to the upper bound $\max(t_1, t_2)$. Since t_1, t_2 are exactly the quantities we would like to estimate, the upper bound on m is however intrinsically unavailable. To resolve this dilemma, we propose an approach with iterative window size change. We obtain estimates of t_1 for several different values of m and then choose one that appears in the sequence of estimates a predefined number of times $n_{\text{oc}} > 1$. The proposed approach is sketched in Algorithm 3. We begin with $m = \lfloor T/2 \rfloor$ and a single value $t_1 = 0$ in the sequence of estimates $\mathcal{T}$ (line 1). We then proceed to obtain the window rank estimates $r_p, p = 1, \ldots, T - m + 1$ (lines 4-6), based on which we append $\mathcal{T}$ with the current estimate of t_1 computed according to (4.25) (line 7). In the last step in lines 8-11, we update the window size by subtracting a fixed δ from it. The update step δ is reduced (line 10) if previous

Algorithm 3: Stationarity interval estimation with automatic window size selection

Input: $\boldsymbol{Y}$, K, n_{oc}, δ

1 Initialize: $\mathcal{T} = \{t_1 = 0\}, m = \lfloor T/2 \rfloor, m_{\text{it}} = 0$

2 **while** $\nexists t_i \in \mathcal{T} : \sum_{k=1}^{|\mathcal{T}|} [t_k = t_i] = n_{\text{oc}}$ **do**

3 $\quad m_{\text{it}} = m_{\text{it}} + 1$;

4 $\quad$ **for** $1 \leq p \leq T - m + 1$ **do**

5 $\quad\quad \boldsymbol{Y}_p^{\text{w}} = [\boldsymbol{y}_p, \ldots, \boldsymbol{y}_{p+m-1}]$;

6 $\quad\quad \hat{r}_p = \text{rank}\left(\boldsymbol{Y}_p^{\text{w}}\right)$;

7 $\quad \mathcal{T} = \{\mathcal{T}, p_{\text{fc}} + m - 1, p_{\text{lc}} - 1\}$;

8 $\quad m = m - \delta$;

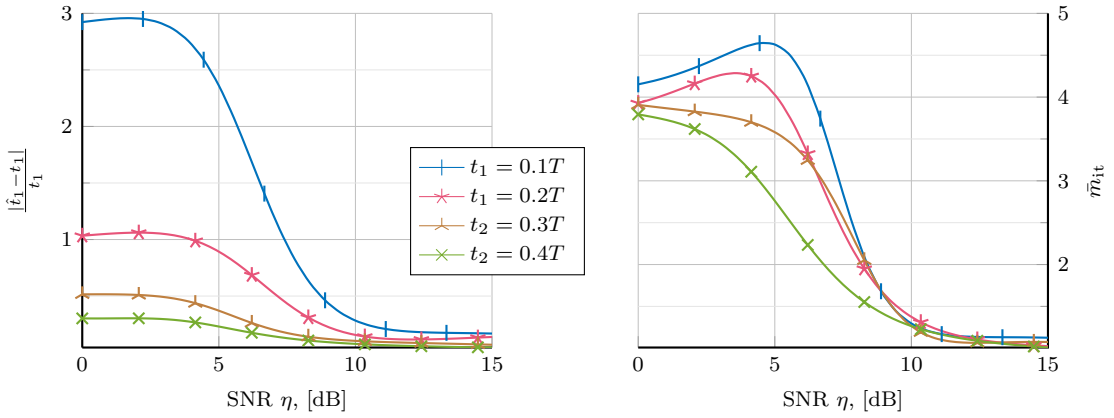
9 $\quad$ **if** $m < K + 1$ **then**

10 $\quad\quad \delta = \min\left(\lfloor \delta/2 \rfloor, 1\right)$;

11 $\quad\quad m = T/2 - \delta$;

12 $\hat{t}_1 = t_i : \sum_{k=1}^{|\mathcal{T}|} [t_k = t_i] = n_{\text{oc}}$;

Output: $\hat{t}_1$



(a) Normalized RMSE **(b)** Number of estimation steps

Figure 4.13: SIE performance evaluation with $n_{\text{oc}} = 2$ and $\delta = 0.1T$: normalized RMSE (a) and number of different window sizes m_{it} (b). Window rank is estimated via ETT.

estimates did not suffice to fulfill the condition on the number of occurrences n_{oc}. Note that the notation $[t_k = t_i]$ in line 2 corresponds to the so-called Iverson bracket notation. Iverson brackets convert any logical proposition placed within them into a number that is 1 if the proposition is satisfied, and 0 otherwise.

4.3.3.2 Numerical evaluation

Figure 4.13 shows the performance of the proposed stationarity interval estimation approach with respect to the RMSE (Figure 4.13a) and the number of different window sizes m_{it} (Figure 4.13b). We consider a quasi-stationary MMV setup with $N = 150$, $M = 15$, $T = 60$ and a single instance of support change where the stationarity interval t_1 varies from $0.1T$ to $0.4T$, while the support difference $\Delta \in \{2k\}_{k=1}^{2K}$ is generated uniformly at random and $K = 5$. The window rank estimates r_p are obtained via the application of the ETT with $\alpha = 10^{-3}$. The results are averaged among $5 \cdot 10^3$ iterations. We see that the estimation performance improves with the SNR and the value of t_1/T. The former is to be expected as a reliable estimation is only possible when the window rank (sparsity order) is estimated correctly. Note that comparing Figure 4.13a with the results for ETT performance presented in Figure 4.3a we see that the SNR ranges for reliable sparsity order and stationarity interval estimation correspond to each other. The fact that an accurate stationarity interval estimation becomes more difficult for lower values of t_1/T is also a likely outcome since the range of indices where the rank of $\boldsymbol{Y}_p^{\text{w}}$ exhibits changes becomes smaller and, as a result, the estimation procedure becomes more susceptible to erroneous rank estimates. Note that increasing t_1 beyond $0.5T$ will reverse the trend since the scenario repeats with respect to $t_2 = T - t_1$. The average number of required estimation steps $\bar{m}_{\text{it}}$ behaves in a similar manner: it decreases with the increase of the SNR and t_1/T from ≈ 5 to the minimal value of 1.

To investigate how the application of the proposed sparsity pattern change detection affects the recovery performance, we perform a following numerical experiment. In a quasi-

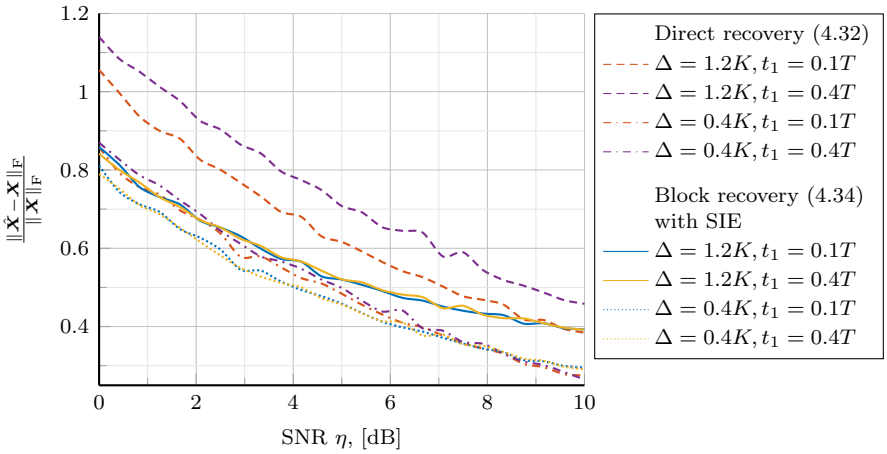


Figure 4.14: Recovery performance evaluation in terms of the normalized RMSE of direct MMV recovery (4.16) and partial MMV recovery (4.18) with SIE where $n_{\mathrm{oc}} = 2$, $K = 5$ and $\delta = 0.1T$.

stationary MMV scenario from Figure 4.13, we recover $\boldsymbol{X}$ by either solving (4.16) directly and treating it as a typical MMV problem, or by applying the SIE approach as described in Algorithm 3 and then solving two independent MMVs according to (4.18). In both cases, for sparse recovery we employ SOMP algorithm where as a stopping criterion we use the true sparsity order: $K + \Delta/2$ for direct recovery according to (4.16) and K for block recovery (4.18). The resulting normalized MSE between the true ($\boldsymbol{X}$) and the estimated ($\hat{\boldsymbol{X}}$) input signals is presented in Figure 4.14. We show the results for two values of the support difference ($\Delta = 1.2K = 6$ and $\Delta = 0.4K = 2$ which correspond to $|\mathcal{S}(\boldsymbol{X}_1) \bigcap \mathcal{S}(\boldsymbol{X}_2)| = 1$ and $|\mathcal{S}(\boldsymbol{X}_1) \bigcap \mathcal{S}(\boldsymbol{X}_2)| = 4$, respectively) and two values of t_1. From these, we see that block sparse recovery with SIE allows to improve estimation performance compared to recovering whole $\boldsymbol{X}$ at once without support change detection. The improvement grows with the increase of the support difference and the relative size of $\boldsymbol{X}_1$ in $\boldsymbol{X}$.

4.3.4 A note on possible model extensions

To conclude, we would like to comment on several cases that do not directly fall within the considered model, but are nevertheless closely related and can therefore be treated in a similar fashion. Cnsider first the case when the individual signals $\boldsymbol{x}_i$ are not necessarily all exactly K-sparse, while $|\mathcal{S}(\boldsymbol{X}_1)| = |\mathcal{S}(\boldsymbol{X}_2)| = K$. In this case, the window rank does not strictly follow (4.24) but rather varies around the values specified by (4.24) and Figure 4.12. This, of course, disrupts the estimation procedure described above as the window rank can potentially change for each value of p. Nevertheless, the average trend will be preserved (i.e., the behavior outlined in Figures 4.12a-c), which can be exploited in order to detect the moment of support change by means of curve fitting for instance. Note that one would expect a performance deterioration in this case compared to the original scenario. This scenario is however of a rather theoretical interest since such an arbitrary appearance and disappearance of active components in the input signal is rarely practically justified.

A somewhat more practical situation is when $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are exactly K_1- and K_2-sparse, respectively, where $\max(K_1, K_2) \leq K$ and $K_1 \neq K_2$. Such a situation can occur, for example, when some signal sources appear on the scene while others disappear from it. It also models the case when one source changes its "position" in the basis space (e.g., a moving target in case of DoA estimation or radar processing, or a frequency hopping transmitter in case of multiband communications). Interestingly, one can show that the behavior of the window rank in this case essentially stays similar to that depicted in Figures 4.12a-c, where the exact type depends on the relation between the supports of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ while suitable range of window sizes reduces from $(K, \max(t_1, t_2)]$ to $(\max(K_1, K_2), \min(t_1, t_2)]$ [A14]. Therefore, the estimation approach in this case will stay largely unchanged.

4.4 Single snapshot sparsity order estimation

Having discussed sparsity order estimation in a particular scenario where the signal scene can be measured multiple times and it provides linearly independent measurements, we turn now to considering a more challenging case where either only a single snapshot is available or the scenario is completely static so that observing it multiple times with the same measurement matrix does not provide linearly independent observations.

In other words, we return to the SMV CS setting where

$$\boldsymbol{y} = \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Psi}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}. \tag{4.26}$$

Moreover, we assume that $\boldsymbol{\Phi}$ can be designed freely and the basis $\boldsymbol{\Psi}$ is an $N \times N$ identity matrix[13]. In the following, we develop a measurement matrix design that recovers linear independence and thus allows estimation of K from the effective rank of a matrix constructed by concatenating blocks of the observed vector along its columns. Note that since $\boldsymbol{A} = \boldsymbol{\Phi}$, we assess our design with respect to $\boldsymbol{A}$ directly.

4.4.1 Measurement design for rank estimation

We begin with considering the noise-free case $\boldsymbol{n} = \boldsymbol{0}$. In order to recover the sparsity order K from $\boldsymbol{y}$ we would like to break $\boldsymbol{y}$ into smaller blocks $\boldsymbol{y}_b \in \mathbb{C}^{m\times 1}$, $b = 1, 2, \ldots, B$ and define a matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_B] \in \mathbb{C}^{m\times B}$ such that $\operatorname{rank}\{\boldsymbol{Y}\} = K$. From (4.26) it is clear that

$$\boldsymbol{y}_b = \boldsymbol{A}_b\boldsymbol{x}, \tag{4.27}$$

where $\boldsymbol{A}_b \in \mathbb{C}^{m\times N}$ contains the m rows that correspond to the bth block of $\boldsymbol{A}$. This sparks the question which condition $\boldsymbol{A}$ must fulfill such that $\operatorname{rank}\{\boldsymbol{Y}\} = K$ for any K-sparse $\boldsymbol{x}$.

[13]Note that as long as the basis $\boldsymbol{\Psi}$ is invertible and known when designing $\boldsymbol{\Phi}$, this holds without loss of generality since for $\boldsymbol{\Psi} \neq \mathbf{I}_N$ we can replace $\boldsymbol{\Phi}$ by $\bar{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$ and achieve the same result.

Theorem 5. *For B non-overlapping blocks of size $m = \frac{M}{B}$, any $K \leq \min(B, m)$ and any K-sparse $\boldsymbol{x}$, we have that* $\operatorname{rank}\{\boldsymbol{Y}\} = K$ *if and only if $\boldsymbol{A} = \boldsymbol{C} \diamond \boldsymbol{A}_0$, where $\boldsymbol{C} \in \mathbb{C}^{B \times N}$ and $\boldsymbol{A}_0 \in \mathbb{C}^{m \times N}$, the Kruskal-rank of $\boldsymbol{C}$ and $\boldsymbol{\Phi}_0$ is $\geq K$, and $\diamond$ denotes the column-wise Kronecker (Khatri-Rao) product.*

Proof. Cf. Appendix B.2.2. ■

It follows from the theorem that the sparsity order must satisfy $K \leq \min(m, B)$. Since $m = \frac{M}{B}$, to maximize this upper bound it is best to choose $m = B = \sqrt{M}$ if M is a square number.

We now move to the case where we let blocks overlap. To this end, we divide $\boldsymbol{y} \in \mathbb{C}^{M \times 1}$ into B blocks of m samples with an offset of p samples from block to block. In other words, the b-th block $\boldsymbol{y}_b \in \mathbb{C}^{m \times 1}$ contains samples $(b-1)p+1$ up to $(b-1)p+m$. To cover all M samples with B blocks we therefore obtain the condition $(B-1)p + m = M$ which implies that the number of blocks is given by $B = \frac{M-m}{p} + 1$ and that $M - m$ must be divisible by p. The case $p = m$ is the one where the blocks do not overlap. For the overlapping case $1 \leq p < m$ the following theorem provides the required structure for $\boldsymbol{A}$ to allow obtaining K from $\boldsymbol{Y}$.

Theorem 6. *For B overlapping blocks of size m (with an offset of p samples between blocks), any $K \leq \min(B, m)$ and any K-sparse $\boldsymbol{x}$, we have that* $\operatorname{rank}\{\boldsymbol{Y}\} = K$ *if and only if the matrix $\boldsymbol{A}$ is constructed by taking the first M rows of $\boldsymbol{C} \diamond \boldsymbol{A}_0$, where $\boldsymbol{A}_0 \in \mathbb{C}^{p \times N}$, $\boldsymbol{C} \in \mathbb{C}^{\left\lceil \frac{M}{p} \right\rceil \times N}$, while $\boldsymbol{C}$ is a Vandermonde matrix, $\boldsymbol{A}$ and $\boldsymbol{C}$ have a Kruskal-rank $\geq K$, and $\lceil \cdot \rceil$ denotes the operation of rounding to the next larger integer.*

Proof. Cf. Appendix B.2.3 ■

Theorem 6 implies that in the special case of $p = 1$ (maximal overlap), the entire sensing matrix $\boldsymbol{A}$ must be a Vandermonde matrix. As briefly discussed in Section 2.2.2, Vandermonde structured measurement matrices have been proposed in the CS context before [220], [221] and rank recovery for Vandermonde mixtures has been studied in the context of harmonic retrieval [222]. In fact, for the case of $p = 1$ the mapping from $\boldsymbol{y}$ to $\boldsymbol{Y}$ is known as spatial smoothing [223] in the harmonic retrieval context where it is applied as a pre-processing step for subspace-based estimators in order to de-correlate coherent signals.

This said, both Theorems 5 and 6 provide conditions on the sensing matrix $\boldsymbol{A}$ in the noise-free setting. When the input signal is contaminated with additive noise according to (4.26), the effective rank of $\boldsymbol{Y}$ has to be estimated, e.g., by applying available MOS algorithms as discussed in the previous sections. From the perspective of SOE, larger block overlap then allows us to increase the number of "snapshots" available for order estimation.

Example 4.15

To demonstrate the single-snapshot sparsity order estimation based on the proposed design, we consider the recovery of a $K = 7$-sparse vector $\boldsymbol{x}$ of length $N = 1000$ from a single vector of observations $\boldsymbol{y}$ of length $M = 256$. We compare three different strategies. Firstly, we set $m = p = 16$ so that there is no overlap and thus $\boldsymbol{\Phi}_0$ and $\boldsymbol{C}$ can be chosen freely. Secondly, we choose $p = 1$ for maximum overlap, in which case $\boldsymbol{A}$ needs to be Vandermonde, where we set $m = 128$ to maximize the size of $\boldsymbol{Y}$. Thirdly, as an intermediate case, $p = 8$ in which case $m = 32$ leads to the largest matrix $\boldsymbol{Y}$ and $\boldsymbol{A}$ is composed of an 8×1000 Vandermonde matrix $\boldsymbol{C}$ and an arbitrary 32×1000 matrix $\boldsymbol{A}_0$. For a direct comparison, the size of the data matrix $\boldsymbol{Y}$ that is used for rank estimation is 16×16 in the non-overlapping case, 128×129 in the case of $p = 1$, and 32×29 for $p = 8$. The matrices that we can choose freely ($\boldsymbol{C}$ in the case of no overlap and $\boldsymbol{A}_0$ in all cases) are drawn from a zero-mean complex Gaussian distribution. The K non-zeros in the vector $\boldsymbol{x}$ are placed uniformly at random and their values are given by $\mathrm{e}^{\jmath\varphi_k}$ where $\varphi_k \sim \mathcal{U}[0, 2\pi)$. The noise $\boldsymbol{n}$ is also drawn from a zero-mean complex Gaussian distribution with variance σ_n^2, while the SNR is defined as $\mathrm{SNR} = 1/\sigma_\mathrm{n}^2$. The sparsity order is estimated by applying existing model order selection criteria to the singular values of the matrix $\boldsymbol{Y} \in \mathbb{C}^{m \times B}$. In particular, we consider AIC [210] as well as the ETT described in Section 4.2.2.2.

Figure 4.15 shows the estimated model order as a function of the SNR, averaged over 1500 Monte-Carlo trials. It demonstrates that the correct sparsity order can be identified from $\boldsymbol{y}$, provided the SNR is not too low. AIC suffers from the very small sample support (which is assumed to be large in the original derivation), which is handled much better by the ETT. We also notice that a larger amount of block overlap (a smaller value of p) provides a better sparsity order estimation performance. This is not surprising since

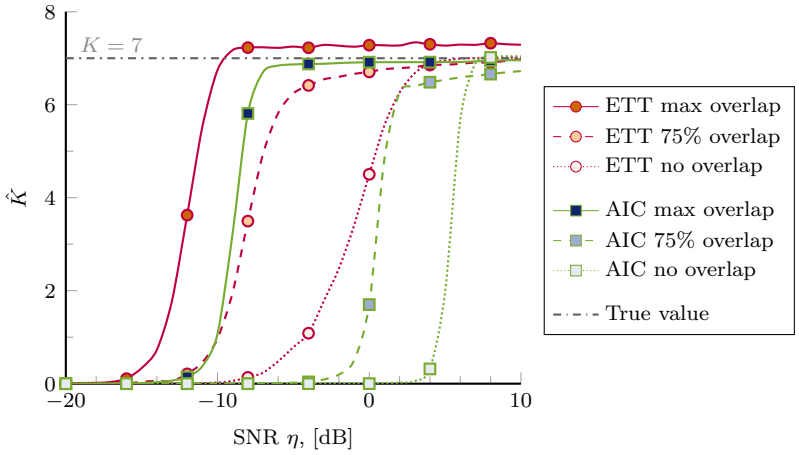


Figure 4.15: Estimated sparsity order $\hat{K}$ vs. SNR for $M = 256$ and $K = 7$. We compare three cases: $p = 1, m = 128$ ("max overlap"), $p = 8, m = 32$ ("75 % overlap") and $p = m = 8$ ("no overlap"). The constant line $K = 7$ shows the true value of the sparsity order.

the size of the matrix $\boldsymbol{Y}$ used for the rank estimation step grows with increasing block overlap. We also observe that ETT tends to slightly overestimate the model order for the case of maximum overlap.

Example 4.15 confirms that a larger block overlap is preferable for SOE. However, it implies an increased level of structure in the sensing matrix $\boldsymbol{A}$, which, as we know, is generally undesirable from the recovery point of view. Below we analyze the proposed design in terms of the coherence of the sensing matrices that are obtainable with the derived structures.

4.4.2 Coherence analysis

For simplicity, let us assume that M is divisible by p which allows us to write $\boldsymbol{A} = \boldsymbol{C} \diamond \boldsymbol{A}_0$ for both cases, the only difference being that in the case of overlap $\boldsymbol{C}$ must be Vandermonde whereas without overlap, $\boldsymbol{C}$ can be arbitrary. The coherence of $\boldsymbol{A}$ is defined by (2.16) as

$$\mu(\boldsymbol{A}) = \max_{n_1 \neq n_2 \in [1,2,\ldots,N]} \left| \frac{\boldsymbol{a}_{n_1}^{\mathrm{H}} \boldsymbol{a}_{n_2}}{\|\boldsymbol{a}_{n_1}\|_2 \, \|\boldsymbol{a}_{n_1}\|_2} \right|. \tag{4.28}$$

Using the column-wise Kronecker structure and the fact that $(\boldsymbol{a} \otimes \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{c} \otimes \boldsymbol{d}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{c}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{d}$ as well as $\|\boldsymbol{a} \otimes \boldsymbol{b}\|_2 = \|\boldsymbol{a}\|_2 \, \|\boldsymbol{b}\|_2$ we obtain

$$\begin{aligned}
\mu(\boldsymbol{A}) &= \max_{n_1 \neq n_2 \in [1,2,\ldots,N]} \underbrace{\left| \frac{\boldsymbol{c}_{n_1}^{\mathrm{H}} \boldsymbol{c}_{n_2}}{\|\boldsymbol{c}_{n_1}\|_2 \, \|\boldsymbol{c}_{n_1}\|_2} \right|}_{\mu_{n_1,n_2}(\boldsymbol{C})} \cdot \underbrace{\left| \frac{\boldsymbol{a}_{0,n_1}^{\mathrm{H}} \boldsymbol{a}_{0,n_2}}{\|\boldsymbol{a}_{0,n_1}\|_2 \, \|\boldsymbol{a}_{0,n_1}\|_2} \right|}_{\mu_{n_1,n_2}(\boldsymbol{A}_0)} \\
&= \max_{n_1 \neq n_2 \in [1,2,\ldots,N]} \mu_{n_1,n_2}(\boldsymbol{C}) \mu_{n_1,n_2}(\boldsymbol{A}_0) && (4.29) \\
&\leq \mu(\boldsymbol{C}) \mu(\boldsymbol{A}_0) && (4.30)
\end{aligned}$$

Note that (4.29) is reminiscent of an argument made in [224] (Lemma 1). There, the authors had considered a Kronecker product structure for the sensing matrix which leads to the product bound (4.30) being tight. However, since we consider a Khatri-Rao product, this is not the case here (by permuting the columns of the factors appropriately, a lower value can be achieved). To further analyze (4.29) we note that according to (2.19) for an arbitrary $P \times Q$ matrix, its coherence is bounded by the Welch bound, i.e., $\mu \geq \sqrt{\frac{Q-P}{P(Q-1)}}$. In the case of no overlap ($p = m = \sqrt{M}$), $\boldsymbol{C}$ and $\boldsymbol{A}_0$ can be designed freely. Therefore, if $\boldsymbol{C}$ and $\boldsymbol{A}_0$ achieved the Welch bound, we would have

$$\mu(\boldsymbol{A}) = \sqrt{\frac{N-m}{m(N-1)}} \sqrt{\frac{N-\frac{M}{m}}{\frac{M}{m}(N-1)}} = \frac{N-m}{m(N-1)}. \tag{4.31}$$

However, the Welch bound is not achievable for all matrix dimensions P and Q. In particular, it is known that equiangular tight framess (ETFs) do not exist for $P < Q^2$, a condition which

is satisfied for both $\boldsymbol{C}$ and $\boldsymbol{A}_0$ for $M = m^2$. Empirically, this leads to a higher coherence than what (4.31) predicts.

Example 4.16

Figure 4.16 demonstrates the dependence of the achievable coherence on the block overlap. We consider a scenario where $M = 256$ and $N = 1000$ and compare the case of maximal block overlap ($p = 1$), no overlap ($p = \sqrt{M} = 16$), the intermediate case $p = 8$ and as a reference the case where $\boldsymbol{A}$ is unstructured matrix (which does not allow any sparsity order estimation). For each case, we display the value of the coherence that is predicted by the Welch bound (according to (4.31) and (4.33)) as well as the coherence achieved by drawing the matrices randomly (from a complex Gaussian distribution), choosing the best among $2 \cdot 10^5$ trials. As expected, the coherence we achieve practically is higher than what the Welch bound predicts. However, both the theoretical and the empirical results show the same trend, namely, the larger the block overlap (i.e., the smaller p), the higher the coherence.

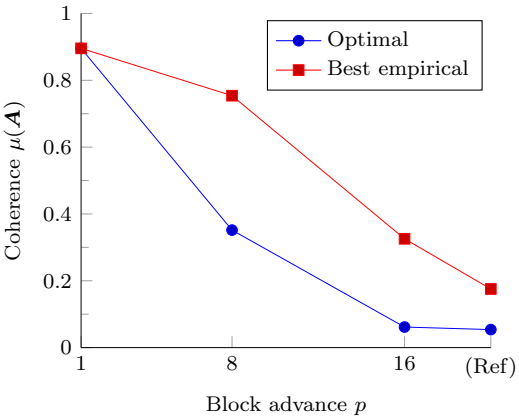


Figure 4.16: Coherence of the matrix $\boldsymbol{A}$ for different amount of block advance: $p = 1$ corresponds to the Vandermonde matrix, $p = 8$ to overlapping blocks, $p = 16$ to non-overlapping blocks, and "(Ref)" to an unstructured matrix $\boldsymbol{A}$.

In the overlapping case, the matrix $\boldsymbol{C}$ must be Vandermonde. Among the class of $P \times Q$ Vandermonde matrices $\boldsymbol{V}$, the frame obtained by considering the top P rows of a $Q \times Q$ DFT matrix with unit-modulus generators [201] achieves the lowest coherence (cf. [225], which also shows that for prime Q these frames have maximal spark). Its coherence is given by considering the inner product between two adjacent columns which can easily be shown to be

$$\mu(\boldsymbol{V}) = \left| \frac{\boldsymbol{v}_n^{\mathrm{H}} \boldsymbol{v}_{n+1}}{\|\boldsymbol{v}_n\|_2 \, \|\boldsymbol{v}_{n+1}\|_2} \right| = \frac{\sin\left(\frac{\pi}{Q} P\right)}{P \sin\left(\frac{\pi}{Q}\right)}. \tag{4.32}$$

Using a similar reasoning as in (4.31), an optimally designed matrix $\boldsymbol{A}_0 \in \mathbb{C}^{p \times N}$ together with the Vandermonde matrix $\boldsymbol{C} \in \mathbb{C}^{\frac{M}{p} \times N}$ achieves a coherence given by

$$\mu(\boldsymbol{A}) = \frac{\sin\left(\frac{\pi}{N} \frac{M}{p}\right)}{\frac{M}{p} \sin\left(\frac{\pi}{N}\right)} \sqrt{\frac{N-p}{p(N-1)}}. \tag{4.33}$$

4.4.3 Choice of the parameters

Theorems 5 and 6 provide an exhaustive answer to the question which sensing strategies allow estimating the sparsity order from the matrix rank of a rearranged version of the single observation vector $\boldsymbol{y}$. As we have shown, there are essentially two parameters we can choose: the block length m and the block advance p. In this section, we discuss the implications of these parameters on the design of $\boldsymbol{A}$ as well as the sparsity order estimation step.

Let us begin with the block length m. Note that the design of the sensing matrix $\boldsymbol{A}$ does not depend on m. Therefore, once a suitable value of p has been selected, the block length m can be chosen without affecting the sensing or the sparse recovery stage. It determines the dimensions of the matrix $\boldsymbol{Y}$ which is important for the rank estimation step that is used to find the sparsity order. More specifically, $\boldsymbol{Y}$ is of size $m \times B$, where $B = \frac{M-m}{p} + 1$. To maximize the size of $\boldsymbol{Y}$, we can choose p such that $m \approx B$, which leads to $m \approx \frac{M+p}{p+1}$. Therefore, we propose to select m as the integer value closest to $\frac{M+p}{p+1}$ such that $M - m$ divides p.

The second parameter we can adjust is the block advance p, which controls the amount of overlap between adjacent blocks (equal to $m - p$ samples). The smaller p is chosen, the more we reuse elements of $\boldsymbol{y}$ which results in a larger overall matrix $\boldsymbol{Y}$. This has a positive effect on the sparsity order estimation step and it allows estimating larger values of K since $K \leq K_{\max} = \min(m, B) - 1$ where $B = \frac{M-m}{p} + 1$. For maximum overlap ($p = 1$), this bound is maximized and becomes $K_{\max} = \frac{M}{2} - 1$ while in the case of no overlap ($p = m = \sqrt{M}$) we have $K_{\max} = \sqrt{M} - 1$. On the other hand, a larger overlap leads to a reduced flexibility in the sensing matrix design since a growing part of $\boldsymbol{A}$ has to obey the Vandermonde scaling law shown in Theorem 6. As discussed above, this has a negative impact on $\mu(\boldsymbol{A})$. Therefore, there is a fundamental trade-off between the performance of the sparsity order estimation and the performance of the sparse recovery step (see Example 4.17). However, note that we consider a system where the measurement matrix $\boldsymbol{A}$ can be adapted at will. This potentially allows switching between measurement matrices designed for the two different purposes: a "probing" matrix which is optimized for the sparsity order estimation step (with a small value of p), and a measurement matrix, which is optimized to the recovery stage (with a larger value of p).

Example 4.17

In this example, we numerically investigate how the block overlap p influences the recovery performance. Figure 4.17 shows the probability of correct support estimation for the same scenario as in example 4.16, using the OMP algorithm for the sparse recovery stage. In addition to the three scenarios discussed above, the curve labeled "Reference" depicts the case where an unstructured matrix $\boldsymbol{A}$ is used for the measurement (which does not allow the single snapshot SOE). We observe that larger overlaps lead to

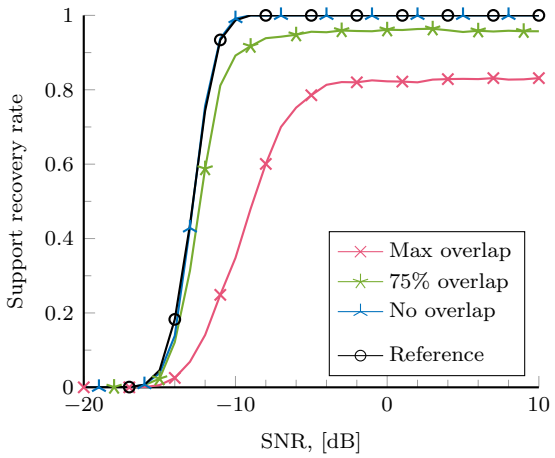


Figure 4.17: Empirical probability of correct support estimation using OMP vs. the SNR for $M = 256$ and $K = 7$. We compare three cases: $p = 1, m = 128$ ("max overlap"), $p = 8, m = 32$ ("75 % overlap") and $p = m = 8$ ("no overlap"). The curve "reference" corresponds to using an unstructured $\boldsymbol{A}$.

a degradation of the recovery performance, which is expected due to the increase in the mutual coherence discussed above. Interestingly, the Khatri-Rao structured matrix $\boldsymbol{A}$ in the non-overlapping case performs almost identically to the unstructured matrix shown as a reference.

4.4.4 A link to multilinear algebra

Beyond enabling the sparsity order estimation in a single snapshot setting, the proposed Khatri-Rao design for the measurement matrix has a strong link to multilinear algebra. In particular, using a Khatri-Rao structured measurement matrix allows rearranging the observed data in a form of a tensor that has (in the noise-free case) a rank-K canonical polyadic decomposition (CPD) [226]. In fact, there are multiple special cases of the proposed designs where such tensors occur, which we would like to list here. For simplicity, let us assume that the parameters p and m are chosen such that p divides M and m.

Firstly, in the case of overlapping blocks we have shown in Theorem 6 that $\boldsymbol{Y} = (\boldsymbol{C}_{\frac{m}{p}} \diamond \boldsymbol{A}_0)\mathrm{diag}\{\boldsymbol{x}\}\boldsymbol{C}_B^{\mathrm{T}} \in \mathbb{C}^{m \times B}$. This matrix can be reshaped into an $m \times p \times B$ tensor $\boldsymbol{\mathcal{Y}}$ which obeys

$$\begin{aligned}\boldsymbol{\mathcal{Y}} &= \boldsymbol{\mathcal{I}}_{3,N} \times_1 \boldsymbol{C}_{\frac{m}{p}} \times_2 \boldsymbol{A}_0 \times_3 (\boldsymbol{C}_B \mathrm{diag}\{\boldsymbol{x}\}) \\ &= \boldsymbol{\mathcal{I}}_{3,K} \times_1 \boldsymbol{C}_{\frac{m}{p},K} \times_2 \boldsymbol{A}_{0,K} \times_3 (\boldsymbol{C}_{B,K} \mathrm{diag}\{\boldsymbol{x}_K\})\,. \end{aligned} \tag{4.34}$$

where $\boldsymbol{\mathcal{I}}_{3,p}$ is the $p \times p \times p$ identity tensor [227], [228]. Moreover, $\boldsymbol{C}_{\frac{m}{p},K}$, $\boldsymbol{A}_{0,K}$, and $\boldsymbol{C}_{B,K}$ contain only the K columns corresponding to the support (i.e., the non-zero elements of of $\boldsymbol{x}$). Obviously, (4.34) is a rank-K CPD which shows that $\boldsymbol{\mathcal{Y}}$ is rank K.

Secondly, in a case where we do have $T > 1$ linearly independent snapshots according to $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$, where $\boldsymbol{Y} \in \mathbb{C}^{M \times T}$, $\boldsymbol{X} \in \mathbb{C}^{N \times T}$ is row-sparse and $\boldsymbol{A} = \boldsymbol{C} \diamond \boldsymbol{A}_0$, applying the

proposed measurement matrix design allows reshaping the given $M \times T$ observation matrix $\boldsymbol{Y}$ into an $B \times p \times T$ tensor $\boldsymbol{\mathcal{Y}}$ where $B = \frac{M}{p}$. Note that this corresponds to the operation we perform to $\boldsymbol{y}$ for non-overlapping blocks, i.e., $m = p$. The resulting tensor can be expressed as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{I}}_{3,N} \times_1 \boldsymbol{C} \times_2 \boldsymbol{A}_0 \times_3 \boldsymbol{X}^{\mathrm{T}} = \boldsymbol{\mathcal{I}}_{3,K} \times_1 \boldsymbol{C}_K \times_2 \boldsymbol{A}_{0,K} \times_3 \boldsymbol{X}_K^{\mathrm{T}},$$

where, as above, $\boldsymbol{X}_K^{\mathrm{T}}$ contains only the columns of $\boldsymbol{X}^{\mathrm{T}}$ corresponding to the support (i.e., the non-zero rows of $\boldsymbol{X}$).

Thirdly, we can combine these two approaches for the case of multiple snapshots and overlapping signals. In this case we transform the $M \times T$ observations into an $m \times p \times B \times T$ tensor with loading matrices given by $\boldsymbol{C}_{\frac{m}{p},K}$, $\boldsymbol{A}_{0,K}$, $\boldsymbol{C}_{B,K}$, and $\boldsymbol{X}_K^{\mathrm{T}}$. Finally, we could also decompose $\boldsymbol{A}$ into more than two matrices, e.g., $\boldsymbol{A} = \boldsymbol{C}_1 \diamond \ldots \diamond \boldsymbol{C}_G \diamond \boldsymbol{A}_0$ which can then be reshaped into a $(G+1)$-dimensional tensor if we have a single snapshot and into a $(G+2)$-dimensional tensor in the multiple snapshot case.

Exploring the potential benefit of this rich multilinear structure in our data is an aspect of future work. We see potential benefit of it in an enhanced denoising, similar to the improvement tensor-based subspace estimation schemes have brought for high-resolution parameter estimation [229]. Moreover, the tensor structure can be used to improve the sparsity order estimation step using tensor-based model order estimation (based on, e.g., [230]). Finally, in the case when the sensing matrix is (partially) unknown (e.g., in a distributed setting), a CPD of the observed tensor $\boldsymbol{\mathcal{Y}}$ could be computed to reveal it. Note that connections between CS and tensors have been discussed in slightly different contexts before, e.g., big low-rank tensors with unknown factors [231] or specially structured "block-sparse" tensors [232].

4.5 Concluding remarks

In this chapter, we examined the problem of sparsity order estimation (SOE) from compressed measurements in MMV and SMV settings. By deriving equivalent system models, we showed that in both cases the sparsity order can be estimated as the rank of a properly arranged matrix of (compressive) measurements. In the case of MMV, this requires the availability of linearly independent snapshots, whereas in the SMV setting (when only a single snapshot is available) we can achieve this goal by a specific (Khatri-Rao) design of the sensing matrix. Since the latter brings additional structure into the sensing matrix, we analyzed the resulting matrix coherence and numerically showed the trade-off between the achievable estimation and recovery performance. It is worth noting that the proposed Khatri-Rao design can also be applied to subsequent CS measurements in a setting where we can change the measurement matrix from one measurement to the next. In this case it recovers the required rank for scenarios where the scene is static and would hence not provide linearly independent

observations by itself. It also introduces a number of possible multilinear structures into the data, which could possibly be exploited by applying tensor-based signal processing.

In the presence of additive noise, the effective rank of the matrix of measurements $\boldsymbol{Y}$ cannot be accessed directly but has to be estimated, e.g., by applying model order selection (MOS) techniques. As an alternative, we proposed the empirical threshold test (ETT) which exploits the empirical distribution of the noise eigenvalues obtained during a training/calibration period. It allows us to account for the influence of the measurement process (the sensing matrix, number of snapshots, etc.) on the noise statistics as well as to control the desired false alarm rate. Numerical evaluations of the proposed algorithm reveal its universal applicability in the considered context. On the contrary, other tested state of-the-art MOS schemes perform better or worse depending on a particular scenario (single vs. multiple snapshot, multiband vs. generic setup, measurement vs. signal noise, etc.).

Considering the MMV model, we discussed the stationary scenario where the signal support stays constant within the entire sensing duration and the quasi-stationary case in which the support exhibits changes once in a while. We proposed an approach for support change detection that exploits rank information obtained from a windowed portion of measurement data. Particularly, we analyzed the conditions on the sensing matrix and the window size that ensure that the support change is reflected in the change of the window rank and devised an iterative method to estimate the moment of this change. The proposed approach permits adapting the number of snapshots used for recovery in order to efficiently utilize the joint-sparse structure of MMV.

In conclusion, we would like to point out that although SOE in the compressed domain has the advantage that no additional signal reconstruction is necessary, its main disadvantage is that it places certain requirements to the signal model: availability of linearly independent snapshots in case of MMV and a particular sensing matrix structure in case of SMV. However, we show that when these are satisfied, eigenvalue-based SOE provides a viable means for assessing the signal's sparsity from the measurement data.

Part II

Applications in Sub-Nyquist Sampling

Having examined the foundations of the finite-dimensional CS framework, we now turn to investigating its application in sub-Nyquist sampling and processing of analog continuous signals. Particularly, we consider the tasks of sub-Nyquist acquisition and processing of wideband multiband signals that are of interest to future generation mobile networks, cognitive radio (CR) systems, radar applications and specialized communications. Interestingly, the realization that the minimal sampling rate required for the acquisition of a multiband signal should be defined by the actual occupied bandwidth rather than by the highest possible frequency it can potentially contain has been around for decades. The absence of effective architectures and processing methods implementing this idea, however, halted significant practical developments. The advent of compressed sensing stimulated new research in the area resulting in first receiver prototypes and numerous extensions of the original problem. This being said, in the following chapters we look at multiband sensing from sub-Nyquist samples from three different perspectives. We begin with the classical task of wideband multiband signal acquisition and spectral sensing at a single sensor and propose a coarse compressive energy detector that operates directly on the received low-rate measurements. We then move to a scenario where such a wideband (sub-Nyquist) sensor is equipped with an antenna array which allows us to include the angular domain into the sensing task. We propose the receiver structure, analyze the recovery conditions and develop approaches for angular-spectral sensing from sub-Nyquist samples. At last, we introduce the spatial domain into the sensing task by considering multiple (time-synchronized) sensors distributed over some geographical area that can exchange data between each other or a centralized processing unit. We show that in the sub-Nyquist sampling framework one can reconstruct the relative autocorrelation functions of the signals within the individual sub-bands, which enables further emitter localization.

Chapter 5

Wideband multiband spectrum sensing

5.1 Motivation and related work

The task of acquiring and processing a wide frequency band that is comprised of multiple communication sub-bands is crucial in a number of applications including cognitive radio (CR), spectrum monitoring and radio surveillance. Multiband operation is also of interest for tactical, maritime and underwater communications [25]–[27], as well as for active and passive radar [28], [29] and ultra wideband (UWB) systems [30]–[33]. Furthermore, the concept of multiband communications is closely related to the idea of carrier aggregation [233], [234] which is considered among the candidate transmission techniques for more efficient spectrum usage in future generation of wireless mobile networks. Notwithstanding, multiband signal acquisition is regarded especially important in the context of CR, a concept of flexible spectrum access that has appeared as a response to often ineffective resource management in traditional communication systems with fixed spectrum allocation [235], [236]. It potentially allows exploring and subsequently utilizing available resources across multiple frequency bands and standards [237]–[239], which facilitates the main goal of CR system that is to intelligently adapt its operation in order to opportunistically access currently under-utilized spectrum [240]. Among others, CR systems are often considered particularly suitable for public safety applications [241] and communication deployment in disaster scenarios [A4], [242].

Wideband signal acquisition

The most distinctive feature of multiband operation is that, in contrast to a single-band system, one needs to receive and process a wide frequency band (possibly up to several Gigahertz) that consists of a number of disjoint communication sub-bands. Depending on the particular application scenario, these communication sub-bands can be significantly spaced

from each other, have varied bandwidths and even be allocated to different services and communication standards. Altogether this poses a number of challenges to the design of the RF front-end, data handling and signal processing that are normally not present in more conventional narrowband receiver systems [243]–[246]. By far, the biggest challenge is the need to cope with the wideband requirement at the receiver front-end. One common solution is to apply a scanning strategy that implies employing a tunable narrowband bandpass filter (BPF) that "sweeps" the frequency acquiring a single (narrowband) sub-band at a time [246]–[248]. Clearly, such an approach falls short of the original goal as it does not simultaneously acquire the wideband multiband input. To circumvent this disadvantage it is often proposed to utilize a filterbank of narrowband BPFs that split the wide frequency range of interest into a number of consecutive sub-bands that are received simultaneously [249]–[253]. Such a filterbank can be either implemented at the RF front-end yielding a multi-channel receiver structure [251] or in the digital domain which assumes an availability of a wideband RF front-end. The former has the obvious drawback of increased hardware complexity caused by multiple parallel receiver chains, whose number depends on the frequency range to be covered and the sampling rate in each channel. The latter approach, on the other hand, requires implementing a single-channel wideband front-end capable of simultaneous wideband acquisition which imposes stringent requirements to the hardware components, most notably to the analog-to-digital converters (ADCs) [245]. Finally, another interesting alternative has been proposed in [239] where multiple distributed receivers form a virtual filterbank by sensing a single (different) sub-band each.

Wideband signal acquisition methods discussed so far are based on the traditional Nyquist-rate sampling where the sampling rate is dictated by the instantaneous bandwidth to be sampled. In previous chapters, we saw that the Nyquist-rate requirement can be highly redundant when the signal of interest possesses a sparse representation. In the specific scenario of multiband communications, the sparse structure arises naturally in the frequency domain when only a portion of the available sub-bands is occupied [254]. In this case, one can apply sub-Nyquist sampling paradigm from Section 2.4 in order to acquire the wideband input at a sub-Nyquist rate without loss of information [22], [141], [202]. The multi-coset sampler (MCS) [23], [141], [202] and the modulated wideband converter (MWC) [255] are two particular sub-Nyquist sampling architectures that provide a practical implementation of this idea. Both are multi-channel receiver systems where in each channel the signal is sampled at a low rate with the overall sampling rate (the number of channels times the sampling rate in a single channel) being less than the Nyquist rate. In MCS, the reduction of the sampling rate is performed by using a delay network followed by direct low-rate sampling of the analog wideband input. Well-known disadvantages of this architecture include the need for high-precision delay lines as well as specialized ADCs with high analog bandwidth, as discussed in [147], [255]. In contrast, the MWC first reduces the signal bandwidth by means of analog mixing with periodic sequences followed by a low-pass filter, after which the resulting signal is sampled with a regular low-rate ADC. In both schemes, the obtained

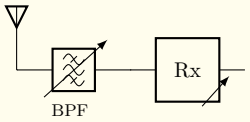
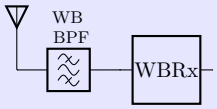
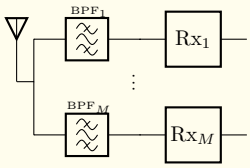
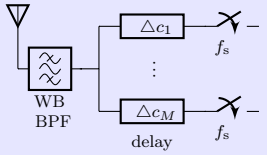
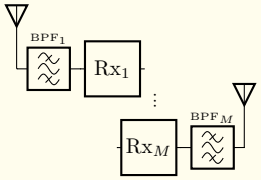
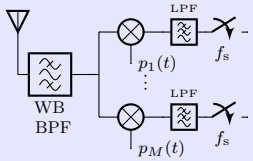
Narrowband RF front-end	Sampling rate f_s		Wideband RF front-end
 **Tune/sweep**	$B \leq f_s \ll 2W$	$f_s \geq 2W$	 **Direct Nyquist**
 **Filter bank**	$f_s \geq B$ $Mf_s \geq 2W$	$f_s \geq B$ $Mf_s \ll 2W$	 **Sub-Nyquist MCS**
 **Distributed**	$f_s \geq B$ $Mf_s \geq 2W$	$f_s \geq B$ $Mf_s \ll 2W$	 **Sub-Nyquist MWC**

Table 5.1: Summary of wideband multiband acquisition methods. In (simplified) circuit diagrams, Rx stands for a narrowband RF chain operating at the bandwidth B of a single frequency sub-band, while WBRx indicates an RF chain that deals directly with the wideband input of bandwidth $W \gg B$. The corresponding sampling rate f_s is given w.r.t B (single sub-band bandwidth) and W (total bandwidth).

low-rate (compressive) samples can then be used to recover the Nyquist-rate signal equivalent, its frequency domain representation or the power spectral density (PSD). Interestingly, it has recently been shown that the sparsity requirement is not necessary for the latter. Namely, the PSD of a wideband input can be estimated from sub-Nyquist samples even for a signal with a densely occupied spectrum, provided that the overall sampling rate is greater than half the Nyquist [256], [257]. The intuition behind this is that the PSD does not contain phase information, which means that its degrees of freedom are twice lower than those of the full frequency spectrum. Furthermore, the PSD is also strictly positive, which further relaxes the recovery conditions. To conclude the discussion on the wideband multiband signal acquisition methods[1] we summarize them in Table 5.1.

Wideband signal processing

After the signal acquisition comes the signal processing step. In this chapter, we consider a particular task of wideband spectrum sensing (WSS) whose aim is to estimate the spectrum occupancy over multiple narrowband communication sub-bands composing the wideband analog input. Spectrum sensing in general and WSS in particular are central to CR, spectrum monitoring and radio surveillance. The task of WSS also finds its place in passive radar

[1]Note that besides MCS and MWC, there are other compressive receiver architectures available in the literature. To mention a few, the so-called random demodulator [258], [259] and the Nyquist folding receiver [260], [261] are the most well-known ones. They are commonly referred to as the analog-to-information (A2I) converters. However, we do not discuss these alternatives here in details, due to their limited applicability.

and sonar applications, as well as in narrowband interference detection [262]. Approaches to WSS have been investigated in a variety of works including techniques based on traditional Nyquist-rate sampling [263]–[273] and methods that make use of sub-Nyquist sampling schemes [256], [274]–[278]. Among these, a seemingly most common WSS approach is to estimate/recover the signal's spectrum from the Nyquist-rate/compressed measurements, split it into multiple narrowband sensing channels and apply conventional signal detection techniques [279], [280], e.g., energy detection, in each of them, either independently [266]–[268], [272] or jointly [269]–[271], [281]. In case of the sub-Nyquist sampled input, such an approach can be applied to the frequency/power spectrum computed from a reconstructed Nyquist sampled signal equivalent or to the PSD estimated directly from the compressed measurements as in [256], [257], [282]. Note that this results in a coarse estimation of the spectrum occupancy since the parameters of the signal sub-bands might not exactly coincide with those of the sensing channels. Often, however, it is enough to provide such an initial coarse evaluation of the spectral occupancy across the observed frequency range. It turns out that to do so one does not necessarily have to perform all the above mentioned steps, i.e., from spectrum estimation to multi-channel narrowband detection. As it was pointed out in [146], the sub-Nyquist sampling framework already provides coarse spectrum sensing functionality at a resolution of the receiver's spectral channelization. In other words, the support of the coefficient vector in the CS formulation indicates which frequency sub-channels are (at least) partially occupied. Since support recovery is an integral part of various efficient signal reconstruction algorithms, such sensing can be often performed automatically as a by-product of the recovery step. In this chapter, we build further on this idea and show that one can perform direct wideband coarse energy sensing from sub-Nyquist samples that can be used without setting additional detection thresholds. Our numerical results suggest that the proposed sensing approach provides best achievable performance compared to the coarse WSS of the same resolution on the recovered PSD, while enjoying lower complexity.

The material of this chapter is partially presented in [A12], [A13], while [A2], [A19] examine some further details on narrowband spectrum sensing and [A4], [A20] address other aspects of CR system design and operation.

5.2 Wideband spectrum sensing

We begin our study with introducing the class of wideband multiband signals and presenting the associated task of wideband spectrum sensing (WSS).

5.2.1 Multiband signal model

Consider a real-valued continuous signal $s(t)$ bandlimited[2] to $\mathcal{F} = [f_{\mathrm{W}} - W/2, f_{\mathrm{W}} + W/2)$ in the positive frequencies where W is the total observed bandwidth and $f_{\mathrm{w}} \geq W/2$ is some central frequency. The energy of $s(t)$ is distributed over K disjoint frequency sub-bands of widths $B_i \leq B$ and central frequencies $f_{\mathrm{c}_i} \in \mathcal{F}$ such that its Fourier transform, given by

$$S(f) = \int_{-\infty}^{+\infty} s(t) e^{-\jmath 2\pi f t} \mathsf{d}t, \tag{5.1}$$

where $S(f) \overset{\Delta}{=} 0,\ f \notin \{\mathcal{F} \cup -\mathcal{F}\}$, is supported on $\mathcal{F}_{\mathrm{s}} = \bigcup_{k=1}^{K} \pm \mathcal{F}_k \subset \mathcal{F}$ with $\mathcal{F}_k = [f_{\mathrm{c}_k} - B_k/2, f_{\mathrm{c}_k} + B_k/2]$. The positions of the signal sub-bands that contain energy are assumed to be unknown beforehand with the individual signals within these sub-bands being independent and uncorrelated. Denoting by $x_k(t)$ the (time domain) signal within the kth active sub-band and by $X_k(f)$ its Fourier transform, we write (5.1) as

$$S(f) = \int_{-\infty}^{+\infty} \sum_{k=1}^{K} x_k(t) e^{-\jmath 2\pi f t} \mathsf{d}t = \sum_{k=1}^{K} \int_{-\infty}^{+\infty} x_k(t) e^{-\jmath 2\pi f t} \mathsf{d}t = \sum_{k=1}^{K} X_k(f), \tag{5.2}$$

where $X_k(f) = 0\ \forall f \notin \pm\mathcal{F}_k$ as illustrated in Figure 5.1.

To each real-valued signal $x_k(t)$ we can associate an equivalent complex-valued (analytic) signal $\dot{x}_k(t) = \frac{1}{2}(x_k(t) + \jmath x_k^{\perp}(t))$ and its (baseband) complex envelope $\bar{x}_k(t) = \dot{x}_k(t) e^{-2\pi f_{\mathrm{c},k} t}$ where $x_k^{\perp}(t)$ denotes the Hilbert transform of $x_k(t)$ [283]. Obviously, we have that $\dot{x}_k(t) = \bar{x}_k(t) e^{2\pi f_{\mathrm{c},k} t}$ and $x_k(t) = 2\mathrm{Re}\{\dot{x}_k(t)\}$ while

$$\dot{X}_k(f) = \begin{cases} X_k(f), & f \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

The complex envelope $\bar{x}_k(t)$, also known as the (complex) baseband representation of $x_k(t)$, as well as the analytic signal $\dot{x}_k(t)$ contain all the information carried by $x_k(t)$ and are hence equivalent representations of the real-valued signal $x_k(t)$ [283].

[2]Note that the bandlimitation condition implies that in practice one is required to apply a pre-selection BPF filter of bandwidth W at the output of the receiving antenna.

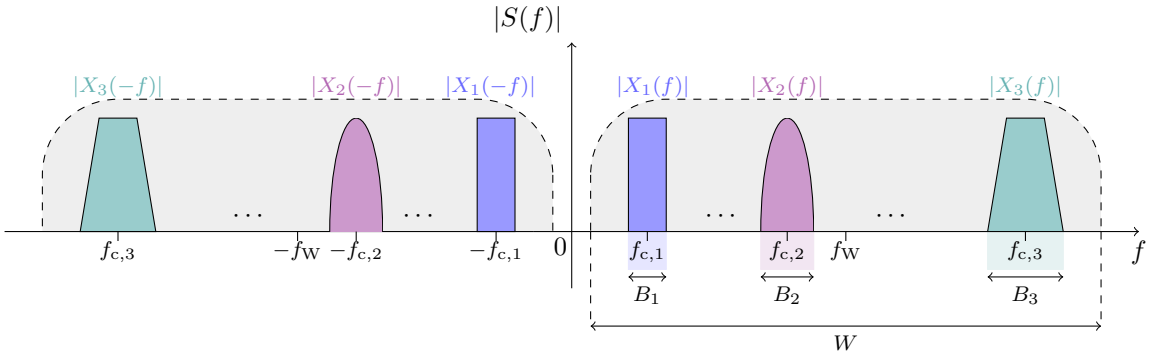


Figure 5.1: An example of a spectrum of a real-valued multiband signal comprised of $K = 3$ active sub-bands located at distinct central frequencies $f_{c,i} \in [f_w - W/2, f_w + W/2]$ with bandwidths $B_i \leq B$.

Given $\dot{x}_k(t)$, we can also represent the real-valued signal $s(t)$ via its complex-valued equivalent[3] $\dot{s}(t) = \sum_{k=1}^{K} \dot{x}_k(t)$ with a Fourier transform $\dot{S}(f) = \begin{cases} S(f), & f \geq 0 \\ 0, & \text{otherwise} \end{cases}$. An example of a spectrum of an analytic signal corresponding to the real-valued signal from Figure 5.1 is provided in Figure 5.2.

Since $s(t)$ and $\dot{s}(t)$ are equivalent, from now on we will use the latter, if not specified otherwise. Furthermore, to simplify the presentation we drop the dot notation of the analytic signal while understanding that $s(t)$ now denotes the complex-valued (analytic) signal equivalent. To conclude, we would like to point out that although f_w can originally be arbitrarily high, its actual value is not crucial for further discussion since for the purpose of signal acquisition the entire band W can be first shifted (possibly in a quadrature way) to some arbitrary chosen IF and only then sampled. Hence, for the sake of brevity we assume henceforth that $f_w = W/2$.

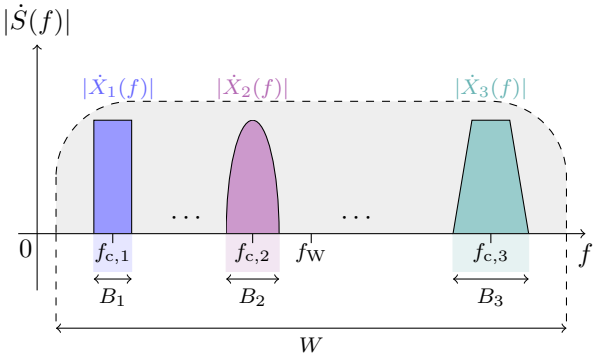


Figure 5.2: An example of a spectrum of an analytic (complex-valued) representation of $s(t)$.

5.2.2 Multiband spectrum sensing

Given an (analytic) wideband input signal $s(t)$ with a multiband spectrum (5.2) (see Figure 5.2), the task of wideband spectrum sensing (WSS) is to determine which parts of

[3]Similarly to $\bar{x}_k(t)$, we could also introduce a complex envelope $\bar{s}(t) = \dot{s}(t)e^{-2\pi f_{sh}t}$. Although the choice of the shift frequency f_{sh} in this case is strictly speaking arbitrary, it makes sense to select it analogously to the central frequencies for the signal sub-bands, e.g., in the middle of the passband W such that it equals f_w. Note that in this case the individual signals $x_k(t)$ are represented by complex envelopes $\tilde{X}_k(f)$ that are shifted from the individual complex envelopes $\bar{X}_k(f)$ by $f_w - f_{c,k}$.

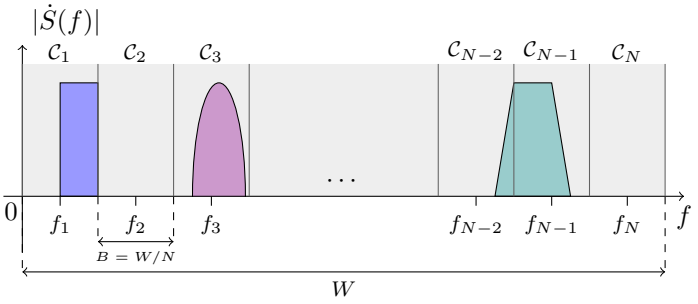


Figure 5.3: Wideband frequency W split into N non-overlapping sensing channels $\mathcal{C}_n$ with central frequencies f_n, $n = 1, 2, \ldots, N$.

the spectrum are occupied and which are not. Available approaches to WSS can be broadly divided into two groups: coarse methods that split the frequency range of interest into a number of non-overlapping sensing channels and apply conventional narrowband spectrum sensing techniques in each of them and fine sensing methods that attempt to detect the edges of the (occupied) signal sub-bands within the wideband frequency range of interest. In this work, we focus on the former and consider coarse wideband spectrum sensing from low-rate sub-Nyquist samples.

5.2.2.1 Coarse sensing

Consider wideband frequency band W being split into N consecutive non-overlapping sensing channels $\mathcal{C}_n = [f_n - B/2, f_n + B/2)$ of width $B = \frac{W}{N} \geq \max_k B_k$, as schematically shown in Figure 5.3. For each $\mathcal{C}_n$ we can formulate a binary hypothesis test

$$\begin{cases} \mathcal{H}_{1,n}: & \mathcal{C}_n \cap \mathcal{F}_\mathrm{s} \neq \emptyset \\ \mathcal{H}_{0,n}: & \text{otherwise} \end{cases}, \tag{5.3}$$

$n = 1, 2, \ldots, N$. To differentiate between $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$, a number of available narrowband detection tests can by applied [279], [284], [285]. By the amount of prior knowledge required, they can be roughly categorized into three major groups: energy-, stochastic feature- and matched filtering based. While the matched filter [217] detectors require the knowledge of the exact signal waveform or at least a part of it, energy detection [286] can be applied in a blind scenario. Moreover, in the absence of prior information on the signal parameters, energy detection is known to be the optimal approach for detecting the presence of signal power in additive noise [286]. Thus using energy based detection, the decision rule for (5.3) can be written as

$$T_n \triangleq \int_{f \in \mathcal{C}_n} |S(f)|^2 \mathrm{d}f \underset{\mathcal{H}_{0,n}}{\overset{\mathcal{H}_{1,n}}{\gtrless}} \zeta_n, \tag{5.4}$$

where ζ_n is the decision threshold for the nth channel

Depending on the individual signal bandwidths B_i and their central frequencies $f_{\mathrm{c},i}$, the sensing channels containing signal energy might be only partially occupied (as it is with $\mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_{N-2} - \mathcal{C}_{N-1}$ in Figure 5.3). Furthermore, a situation can also occur in which $\mathcal{F}_k$ is split between two neighboring sensing channels (e.g., see the third signal sub-band in Figure 5.3 being split between $\mathcal{C}_{N-1}$ and $\mathcal{C}_{N-2}$). This means that an application of the detection strategy (5.3) results in the coarse estimate of the spectral support, i.e., $\mathcal{F}_{\mathrm{s}}^{\mathrm{c}} = \bigcup_{\{q:\mathcal{H}_{1,q} \text{ is true}\}} \mathcal{C}_q \supseteq \mathcal{F}_{\mathrm{s}}$. When $\mathcal{F}_{\mathrm{s}}^{\mathrm{c}}$ is estimated correctly, then the difference between the actual ($\mathcal{F}_{\mathrm{s}}$) and the coarse ($\mathcal{F}_{\mathrm{s}}^{\mathrm{c}}$) spectral supports is

$$0 \leq \delta_f \triangleq \lambda^* \left(\mathcal{F}_{\mathrm{s}}^{\mathrm{c}} \ominus \mathcal{F}_{\mathrm{s}} \right) = \lambda^*(\mathcal{F}_{\mathrm{s}}^{\mathrm{c}}) - \lambda^*(\mathcal{F}_{\mathrm{s}}) \leq KB, \tag{5.5}$$

where $\mathcal{A} \ominus \mathcal{B} = (\mathcal{A} \cup \mathcal{B}) \setminus (\mathcal{A} \cap \mathcal{B})$ stands for the symmetric difference between $\mathcal{A}$ and $\mathcal{B}$ while $\lambda^*(\cdot)$ denotes the Lebesgue measure of an interval. The upper bound in (5.5) is due to the fact that a single signal sub-band can potentially contribute to up to two channels $\mathcal{C}_n$. The actual value of δ_f depends on the values of $f_{\mathrm{c},i}$ and B_i: it achieves 0 when $f_{\mathrm{c},i} \in \{f_n\}_{n=1}^N \; \forall i \in [1, K]$ and $B_i = B$, whereas for $B_i \leq B$ we have that $\delta_f = KB - \sum_i B_i$.

Example 5.18

In this example, we demonstrate how the accuracy of the coarse WSS depends on the difference between the parameters (central frequencies f_n and bandwidth B) of the sensing channels and the parameters (central frequencies $f_{\mathrm{c},i}$ and bandwidths B_i) of the signal sub-bands. We do so on the example of a single signal sub-band with a bandwidth $0.1 \leq B_i/B \leq 0.9$ whose (normalized) central frequency $f_{\mathrm{c},i}^{\mathrm{n}} = f_{\mathrm{c},i}/B - \lfloor f_{\mathrm{c},i}/B \rfloor$ varies from 0 to 1 where $\lfloor \cdot \rfloor$ denotes operation of of rounding to the nearest smaller integer.

Figure 5.4 depicts the normalized difference δ_f/B as a function of $f_{\mathrm{c},i}^{\mathrm{n}}$ and B_i/B. We observe that it has two distinct areas: the light-gray colored triangular area in the middle with the center at $f_{\mathrm{c},k}^{\mathrm{n}} = 0.5$, which corresponds to $f_{\mathrm{c},i} \in \{f_n\}_{n=1}^N$, where the estimation error is upper-bounded by B and the area that consists of two dark-gray triangles at the edges of the considered intervals where the estimation error is within $[B, 2B]$. This illustrates why we refer to the multi-channel detection approach presented above as coarse: it is capable of providing spectrum occupancy estimation at the resolution of B only. Furthermore, Figure 5.4 also indicates that if $B_i \simeq B$ and $f_{\mathrm{c,i}} \notin \{f_n\}_{n=1}^N$,

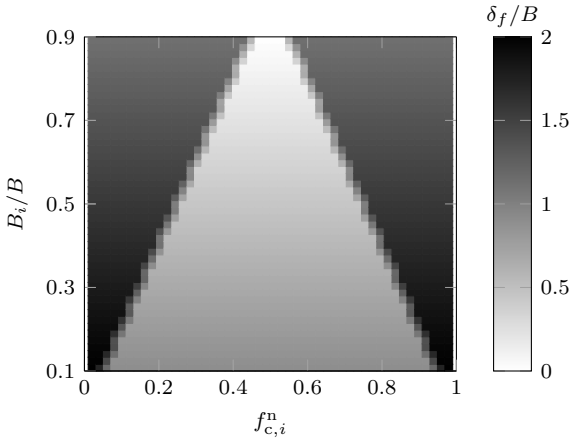


Figure 5.4: Normalized difference δ_f/B vs. normalized signal sub-band position $f_{\mathrm{c},i}^{\mathrm{n}} = f_{\mathrm{c},i}/B - \lfloor f_{\mathrm{c},i}/B \rfloor$ and normalized bandwidth B_i/B.

signal sub-bands are likely to be split between two neighboring sensing channels resulting in $\delta_f \geq B$ and a decrease of the signal-to-noise ratio within each channel. The degree to which this has an effect on the sensing performance is investigated in Section 5.3.3.

5.2.2.2 Performance metrics

Consider an $N \times 1$ binary occupancy vector $\boldsymbol{b}$ and its estimate $\hat{\boldsymbol{b}}$, e.g., obtained by applying the decision rule (5.4), with nth elements given by

$$b_n = \begin{cases} 1, & \text{if } \mathcal{C}_n \cap \mathcal{F}_\text{s} \neq \emptyset \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{b}_n = \begin{cases} 1, & \text{if } T_n \geq \zeta_n \\ 0, & \text{otherwise} \end{cases}, \tag{5.6}$$

respectively. In the spirit of previous discussions, we can introduce the support $\mathcal{S}(\boldsymbol{b})$ of $\boldsymbol{b}$ that contains the indices of the active sensing channels $\mathcal{C}_n$ and hence defines the coarse frequency support $\mathcal{F}_\text{s}^\text{c}$, i.e., $\lambda^*(\mathcal{F}_\text{s}^\text{c}) = |\mathcal{S}(\boldsymbol{b})|B$. Naturally, for each of N sensing channels $\mathcal{C}_n$ we can also define probabilities of false alarm $P_{\text{fa},n}$ and missed detection $P_{\text{md},n}$ as

$$P_{\text{fa},n} \triangleq \Pr\left[\hat{b}_n = 1 | \mathcal{H}_{0,n}\right] = \Pr\left[\hat{b}_n = 1 | b_n = 0\right], \tag{5.7}$$

$$P_{\text{md},n} \triangleq \Pr\left[\hat{b}_n = 0 | \mathcal{H}_{1,n}\right] = \Pr\left[\hat{b}_n = 0 | b_n = 1\right]. \tag{5.8}$$

If the distributions of the test statistics T_n under $\mathcal{H}_{0,n}$ (no part of the signal power is present in nth sensing channel) and $\mathcal{H}_{1,n}$ (at least some part of the signal power is present in nth sensing channel) are known, one could potentially calculate (5.7), (5.8) exactly, as in [239] for instance.

In order to evaluate the overall performance of the multi-channel detector, however, we need some cumulative performance metrics. While different options have been suggested in the literature, we primarily adopt the following

- total probability of correct detection P_d

$$P_\text{d} \triangleq \Pr\left[\hat{b}_n = \begin{cases} 1, & \forall n \in \mathcal{S}(\boldsymbol{b}) \\ 0, & \text{otherwise} \end{cases}\right]. \tag{5.9}$$

- probability of at least one sensing channel $\mathcal{C}_n$ being falsely detected as occupied

$$P_\text{fa} \triangleq \Pr\left[\exists\, n \notin \mathcal{S}(\boldsymbol{b}) : \hat{b}_n = 1\right]. \tag{5.10}$$

- probability of at least one sensing channel $\mathcal{C}_n$ being falsely detected as non-occupied

$$P_\text{md} \triangleq \Pr\left[\exists\, n \in \mathcal{S}(\boldsymbol{b}) : \hat{b}_n = 0\right]. \tag{5.11}$$

- support recovery rate (SRR), familiar to us from Chapter 3, i.e.,

$$\mathrm{SRR} = \frac{|\mathcal{S}(\hat{\boldsymbol{b}}) \cap \mathcal{S}(\boldsymbol{b})|}{|\mathcal{S}(\boldsymbol{b})|}. \tag{5.12}$$

Note that derivation of general relationship linking (5.9)-(5.12) with (5.7)-(5.8) is non-trivial since one has to consider all possible combinations of decision outcomes in different channels. Moreover, the test statistics T_n can be correlated, which further complicates the analysis.

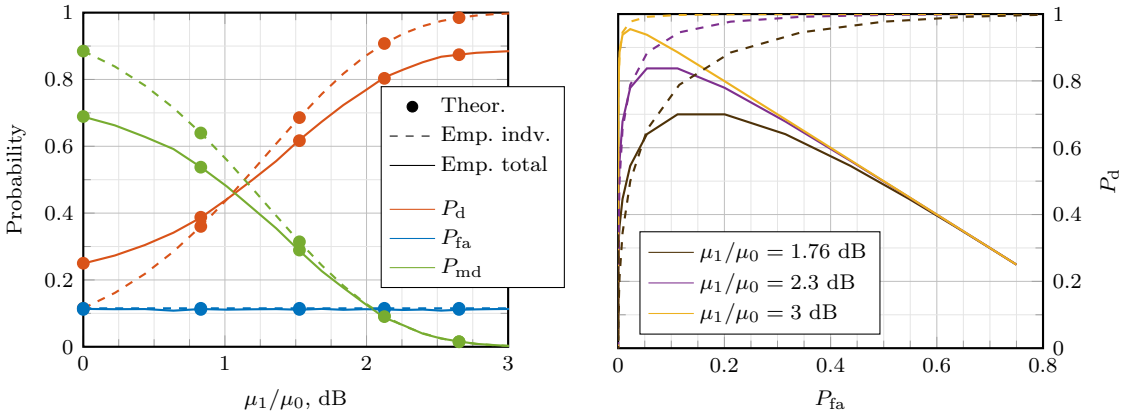
Example 5.19

To appreciate the difference between the individual (per channel) probabilities (5.7)-(5.8) and the total ones (5.9)-(5.11) consider a following toy example. Suppose we have $N = 2$ channels $\mathcal{C}_n$ whose test statistics T_n are independent from each other and follow a normal Gaussian distribution both under $\mathcal{H}_{1,n}$ and under $\mathcal{H}_{0,n}$: $T_n|\mathcal{H}_{0,n} \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right)$ and $T_n|\mathcal{H}_{1,n} \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ for both n. Then, it is easy to show that

$$P_{\mathrm{fa},n} = Q\left(\frac{\zeta_n - \mu_0}{\sigma_0}\right) \text{ and } P_{\mathrm{md},n} = 1 - Q\left(\frac{\zeta_n - \mu_1}{\sigma_1}\right),$$

where $Q(\cdot)$ is the Q-function. Choosing equal thresholds $\zeta_n = \zeta$ yields equal $P_{\mathrm{fa},n}$ and equal $P_{\mathrm{md},n}$, which we denote by α and β, respectively. For a binary occupancy vector $\boldsymbol{b}$ with elements distributed as i.i.d. Bernoulli random variables with $p = 0.5$, meaning that each channel can be empty or occupied with equal probability, one can show (see Appendix B.3.1) that $P_{\mathrm{d}} = \frac{1}{4}(2 - \alpha - \beta)^2$, $P_{\mathrm{fa}} = \frac{1}{4}\alpha(4 - \alpha)$, and $P_{\mathrm{md}} = \frac{1}{4}\beta(4 - \beta)$.

Figure 5.5a presents theoretical and empirical probabilities of correct and missed detection, and false alarm for the case of a fixed μ_0 and a varying μ_1 with $\sigma_1 = \sigma_0 = 0.25\mu_0$ and the threshold $\zeta = 1.3\mu_0$. It shows both the individual probabilities $P_{\mathrm{d},n}$, $P_{\mathrm{fa},n}$, $P_{\mathrm{md},n}$



(a) Probabilities vs. SNR

(b) ROC curves

Figure 5.5: Performance evaluation of the scenario form Example 5.19 with respect to probabilities of false alarm, missed and correct detection (a) and the ROC curves (b).

and the total ones P_{d}, P_{fa}, P_{md}. We see that the individual and total probabilities differ in value; thus P_{d} for instance does not reach 1 in the high SNRs. Yet, the most interesting tendency is revealed in Figure 5.5b that shows corresponding ROC curves. While the curves for individual probabilities exhibit typical detection behavior, i.e., the probability of correct detection increases with the probability of false alarm, the curves for the total probabilities demonstrate a substantially different trend. For low P_{fa} values (high threshold values) they largely follow the curves for individual probabilities. At some point, however, the two begin to depart with P_{d} decreasing sharply with the increase of P_{fa}. The reason for this to happen is that lowering of the detection threshold below a certain values results in a rapid increase of P_{fa} since the probability that the noise energy in at least one of the unoccupied sensing channels exceeds ζ becomes significant. The total probability of correct detection P_{d} then accordingly plummets down as a false alarm constitutes an erroneous estimate in the binary occupancy vector. This shows that in a multiband setting there is an optimal detection threshold value that maximizes P_{d} for a given SNR.

5.3 Wideband sensing from sub-Nyquist samples

Having introduced the class of multiband signals and the associated task of coarse wideband spectrum sensing, we now move to considering a particular sub-Nyquist receiver architecture, the MWC. The MWC is a practical sub-Nyquist sampling architecture that allows compressing the wideband multiband input such that the actual signal sampling can be carried out by low-rate commercially available ADCs [39], [147]. It achieves this by incorporating an RF pre-processing step prior to the sampling that performs an analog bandwidth reduction. In contrast, its closest counterpart, the MCS, does not provide such a mechanism and as a result requires high-bandwidth ADCs [147]. Our choice is further motivated by the versatility of the scenarios the MWC can be applied to [34] and the fact that it has already seen first successful prototype implementations [287].

5.3.1 Modulated wideband converter

5.3.1.1 Receiver architecture

In the MWC, the wideband input $s(t)$ is fed into M parallel sampling channels. In each channel, it is first multiplied by some T_{p}-periodic sequence $p_m(t)$, then passed through an low-pass filter (LPF) with a cut-off frequency $f_{\mathrm{s}}/2$ and sampled at a low sampling rate of f_{s} [39], [147], [255], as depicted in Figure 5.6. The total sampling rate in the MWC is $M f_{\mathrm{s}}$, which is assumed to be less than f_{nyq}. Given the receiver structure from Figure 5.6, the parameters to choose are the number of sampling channels M, the sampling frequency f_{s}, the mixing sequence repetition rate $f_{\mathrm{p}} = 1/T_{\mathrm{p}}$, and the mixing sequences $p_m(t)$ themselves. With regard

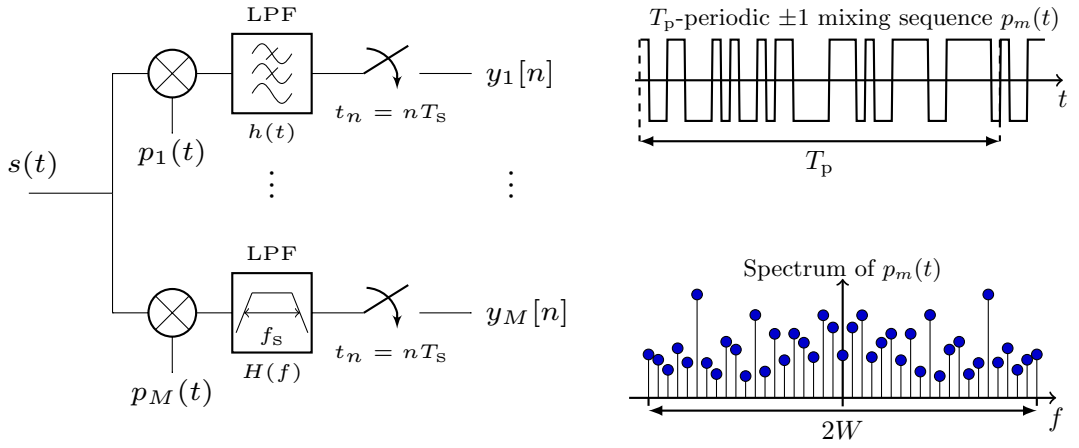


Figure 5.6: Block diagram of the MWC sub-Nyquist sampling architecture with M parallel channels (left-hand side) and an example of a ± 1 pseudo-random mixing sequence $p_m(t)$ in time and frequency domains (right-hand side). Adapted from [255].

to the latter, one generally requires that $p_m(t)$ are T_p-periodic with a bandwidth greater than L/T_p. In [255] for instance, the authors consider (pseudo-random) sign-alternating functions that take value ± 1, as illustrated in Figure 5.6. Importantly, since $p_m(t)$ are periodic they can be represented by a Fourier series as

$$p_m(t) = \sum_{\ell=-\infty}^{\infty} c_{m,\ell} e^{\jmath 2\pi \ell f_p t}, \tag{5.13}$$

where $c_{m,\ell}$ are the Fourier coefficients that are given by

$$c_{m,\ell} = \frac{1}{T_p} \int_0^{T_p} p_m(t) e^{-\jmath 2\pi \ell f_p t} \mathsf{d}t. \tag{5.14}$$

From (5.13)–(5.14) we see that in the frequency domain the mixing sequences $p_m(t)$ present weighted Dirac-combs (see an example in Figure 5.6). Subsequently, its multiplication with $s(t)$ results in spreading of the weighted copies of the individual signal sub-bands over $\mathcal{F}$.

5.3.1.2 Frequency domain analysis

Denote by $\tilde{y}_m(t) = s(t)p_m(t)$ the input signal after mixing with mth sequence $p_m(t)$. Writing down the Fourier transform of $\tilde{y}_m(t)$

$$\begin{aligned} \tilde{Y}_m(f) &= \int_{-\infty}^{\infty} s(t) p_m(t) e^{-\jmath 2\pi f t} \mathsf{d}t = \int_{-\infty}^{\infty} s(t) \left(\sum_{\ell=-\infty}^{\infty} c_{m,\ell} e^{\jmath 2\pi \ell f_p t} \right) e^{-\jmath 2\pi f t} \mathsf{d}t \\ &= \sum_{\ell=-\infty}^{\infty} c_{m,\ell} \int_{-\infty}^{\infty} s(t) e^{-\jmath 2\pi (f - \ell f_p) t} \mathsf{d}t = \sum_{\ell=-\infty}^{\infty} c_{m,\ell} S(f - \ell f_p), \end{aligned} \tag{5.15}$$

we see that the output of mth mixer is a linear combination of f_p-shifted and weighted copies of $S(f)$. Note that since $S(f) = 0, f \notin \mathcal{F}$, it is enough to take $L \leq \lceil f_{nyq}/f_p \rceil + 1$ terms in

the sum in (5.15) [255]. Hence, we can write (5.15) as

$$\tilde{Y}_m(f) = \sum_{\ell=-L_0}^{L_0} c_{m,\ell} S(f - \ell f_\mathrm{p}) = \sum_{k=1}^{K} \tilde{X}_{k,m}(f), \tag{5.16}$$

where $\tilde{X}_{k,m}(f) = \sum_{\ell=-L_0}^{L_0} c_{m,\ell} X_k(f - \ell f_\mathrm{p})$ and $L_0 = \left\lceil \frac{f_\mathrm{nyq}+f_\mathrm{s}}{2f_\mathrm{p}} - 1 \right\rceil$ [255]. This way, at the mixing step the wideband spectrum $\mathcal{F}$ is conceptually divided into spectral cells of width f_p that are shifted by ℓf_p and linearly combined together. A schematic illustration of this process is provided in Figure 5.7.

Assuming that LPFs have ideal rectangular frequency responses $H(f)$, only the part of $\tilde{Y}_m(f)$ that lies in the interval $\mathcal{F}_\mathrm{s} = [-f_\mathrm{s}/2, f_\mathrm{s}/2]$ is preserved after filtering, i.e.

$$\dot{Y}_m(f) = \tilde{Y}_m(f) H(f) = \begin{cases} \tilde{Y}_m(f), & \text{if } f \in \mathcal{F}_\mathrm{s} \\ 0, & \text{otherwise} \end{cases}. \tag{5.17}$$

This way, the mixing and low-pass filtering constitute an RF pre-processing part that reduces the analog bandwidth to be sampled from W to f_s, while ensuring that all signal sub-bands $X_k(f)$ contribute to the signal obtained at the output of the receiver.

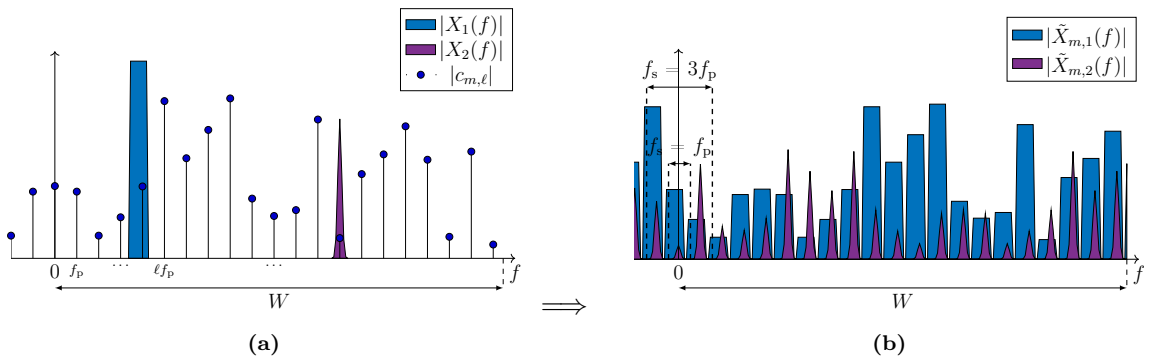
Given (5.17), we can write the DTFT of the (low-rate) output sequence $y_m[n]$ as

$$Y_m(e^{\jmath 2\pi f T_\mathrm{s}}) = \sum_{n=-\infty}^{\infty} y_m[n] e^{-\jmath 2\pi f n T_\mathrm{s}} = \sum_{\ell=-L_0}^{L_0} c_{m,\ell} S(f - \ell f_\mathrm{p}), f \in \mathcal{F}_\mathrm{s}. \tag{5.18}$$

Collecting all $Y_m(e^{\jmath 2\pi f T_\mathrm{s}})$ into one length-M vector $\boldsymbol{y}$ we arrive at the matrix-form relation that closely resembles a finite-dimensional CS system considered in Part I of this thesis, i.e.,

$$\boldsymbol{y}(f) = \boldsymbol{A}\boldsymbol{z}(f), f \in \mathcal{F}_\mathrm{s}. \tag{5.19}$$

The main difference of (5.19) from (2.2) is that it represents an underdetermined system of linear equations over a continuous parameter f: here, the mth element of $\boldsymbol{y}(f)$ is $Y_m(e^{\jmath 2\pi f T_\mathrm{s}})$



(a) **(b)**

Figure 5.7: Spectra of individual signals $x_k(t)$ before (a) and after (b) mixing with a ± 1 pseudo-random ($T_\mathrm{p} = 1/f_\mathrm{p}$)-periodic function $p_m(t)$.

while $\boldsymbol{z}(f)$ is a length-L vector with ℓth element

$$z_\ell(f) = S(f - \ell f_\mathrm{p}), f \in \mathcal{F}_\mathrm{s}. \tag{5.20}$$

The matrix $\boldsymbol{A}$ in (5.19) is an $M \times L$ sensing matrix that contains the Fourier coefficients $c_{m,\ell}$ such that $[\boldsymbol{A}]_{m,\ell} = c_{m,-\ell} = c^*_{m,\ell}$, e,g., for $f_\mathrm{s} = f_\mathrm{p}$ it is composed as

$$\boldsymbol{A} = \begin{pmatrix} c^*_{1,L_0} & c^*_{1,L_0-1} & \cdots & c_{1,0} & \cdots & c_{1,L_0-1} & c_{1,L_0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ c^*_{M,L_0} & c^*_{M,L_0-1} & \cdots & c_{M,0} & \cdots & c_{M,L_0-1} & c_{M,L_0} \end{pmatrix}. \tag{5.21}$$

Note that (5.19) can be equivalently written in time domain by using time samples $\boldsymbol{y}[n] = [y_1[n], \ldots, y_M[n]]^\mathrm{H}$ as

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{z}[n], \tag{5.22}$$

where $\boldsymbol{z}[n]$ is the inverse discrete time Fourier transform (IDTFT) of $\boldsymbol{z}(f)$.

This way, expressions (5.19), (5.20) bind the known DTFT of the receiver output $\boldsymbol{y}[n]$ to the unknown spectrum $S(f)$ of the wideband input $s(t)$ rearranged into $\boldsymbol{z}(f)$.

5.3.1.3 Choice of the parameters

The choice of the parameters for perfect recovery of $S(f)$ from $\boldsymbol{y}[n]$ is amply explained in the original publications [39], [255]. Rather than repeating the full analysis here, we briefly outline the main considerations that governs the MWC settings.

Frequencies f_s and f_p. From (5.17), the repetition rate f_p (or equivalently the sequences period T_p) defines the frequency shifts of $S(f)$ and, as a consequence, the arrangement of the shifted copies of $X_k(f)$ in $\boldsymbol{z}(f)$. In order for each signal sub-band to be represented by a single entry[4] in $\boldsymbol{z}(f)$, referring to any particular frequency $f \in \mathcal{F}_\mathrm{s}$, f_p has to be bigger or equal to $B \geq \max_k B_k$. On the other hand, we also need to ensure that the signal sub-bands are fully represented in $\boldsymbol{z}(f)$, which requires that $f_\mathrm{s} \geq f_\mathrm{p}$ since only those frequencies are contained in the output sequences $y_m[n]$ that fall within $\mathcal{F}_\mathrm{s}$. Therefore, we have that $B \leq f_\mathrm{p} \leq f_\mathrm{s}$, which yields the minimal setting of $f_\mathrm{p} = f_\mathrm{s} \approx B$.

Note that, if the condition on f_p is fulfilled, it is not necessarily required to have $f_\mathrm{s} = f_\mathrm{p}$. One could increase f_s and include several adjacent spectral cells into $\mathcal{F}_\mathrm{s}$, as shown in Figure 5.8. This would allow one to trade the sampling frequency f_s for the number of channels M. To obtain a representation equivalent to (5.19), however, one would then require to (digitally) pre-process $\boldsymbol{y}[n]$ first. We examine this case in the following section.

[4] Failure to do so will result in an inherent ambiguity which cannot be resolved during the later processing.

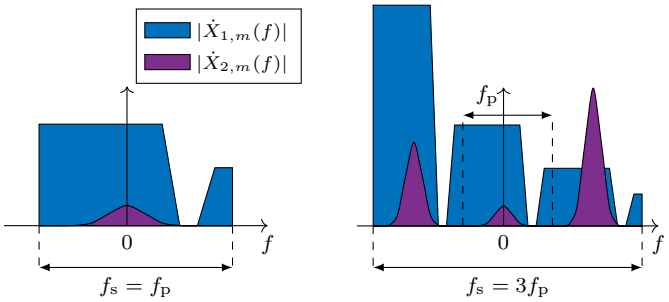


Figure 5.8: Spectra of the individual signals from Figure 5.7 after low-pass filtering with $f_\mathrm{s} = f_\mathrm{p}$ and $f_\mathrm{s} = 3f_\mathrm{p}$ where $f_\mathrm{p} \approx B$.

Mixing sequences $p_m(t)$. We now turn to discussing the choice of the mixing sequences $p_m(t)$. It has been already noted before that $p_m(t)$ have to be T_p-periodic with a bandwidth of at least Lf_p so that their spectrum is discrete with L significant coefficients $c_{m,\ell}$ located at ℓf_p. The mixing sequences also define the sensing matrix $\boldsymbol{A}$ by providing $c_{m,\ell}$, i.e., mth row of $\boldsymbol{A}$ corresponds to mth sequence $p_m(t)$, as is evident from (5.21). Therefore, $p_m(t)$ should be i) different to provide different linear combinations of $X_k(f)$, ii) result in a sensing matrix that ensures recoverability of $X_k(f)$ from $\boldsymbol{y}[n]$, e.g., in terms of the CS conditions stated in Section 2.2. This being said, pseudo-random binary sequences such as Kasami, Gold and m-sequences are considered particularly well suited candidates [288].

Number of sampling channels M. Finally, the number of sampling channels M defines the total sampling rate Mf_s as well as the number of measurements in CS formulation (5.19), (5.22). It has been demonstrated that in the noise-free setting the necessary and sufficient condition for exact signal recovery is $M \geq 2K$.

5.3.1.4 Spectrum recovery

Having determined the relation between the low-rate receiver output $\boldsymbol{y}[n]$ and the wideband input $s(t)$, we now show how one can recover $S(f)$ from $\boldsymbol{y}[n]$, provided that $M \geq 2K$, $f_\mathrm{s} \geq f_\mathrm{s} \geq B$ and $\sum_k B_k \ll W$. Note that recovering of $S(f)$ is tantamount[5] to recovering $\boldsymbol{z}[f]$. Therefore, in the following we limit our discussion to the point of obtaining $\boldsymbol{z}(f)$ from $\boldsymbol{y}[n]$, considering separately: i) recovery under minimal conditions $f_\mathrm{s} = f_\mathrm{p} \approx B$, referred to as the basic configuration, and ii) the expander scheme with $f_\mathrm{s} = (2k+1)f_\mathrm{p}$, $k \geq 1 \in \mathbb{Z}$. Generally, in both cases we assume that $f_\mathrm{p} \approx B$. As a result, each $X_k(f)$ contributes a single non-zero to $\boldsymbol{z}(f)$ for any given $f \in \mathcal{F}_\mathrm{s}$. Therefore, $\boldsymbol{z}(f)$ is maximally K-sparse with respect to any particular frequency $f \in \mathcal{F}_\mathrm{s}$. However, since each sub-band can be split between up to two adjacent spectral cells, the overall sparsity order with respect to the entire frequency range $\mathcal{F}_\mathrm{s}$ is upper bounded by $2K$.

[5] When $\boldsymbol{z}[f]$ is known and the conditions described above are fulfilled, one can obtain $S(f)$ by shifting the spectrum cells $z_\ell(f)$ back by $-\ell f_\mathrm{p}$.

Recovery with $f_\mathrm{s} = f_\mathrm{p}$. We begin by noting that (5.22) and (5.19) correspond to the IMV problem (2.39) that arose in the context of SI union of sub-spaces model discussed in Section 2.4. We can use therefore the continuous-to-finite (CTF) block [137] to reduce it to the MMV recovery problem, as exemplified by (2.40)–(2.41). To do so, we first compute from $\boldsymbol{y}[n]$ an observation covariance matrix $\boldsymbol{R}$ as

$$\boldsymbol{R} = \sum_{n=-\infty}^{\infty} \boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n] = \int_{f\in\mathcal{F}_\mathrm{s}} \boldsymbol{y}(f)\boldsymbol{y}^{\mathrm{H}}(f)\mathsf{d}f. \tag{5.23}$$

Substituting (5.19) into (5.23), we obtain

$$\boldsymbol{R} = \int_{f\in\mathcal{F}_\mathrm{s}} \boldsymbol{A}\boldsymbol{z}(f)\boldsymbol{z}^{\mathrm{H}}(f)\boldsymbol{A}^{\mathrm{H}}\mathsf{d}f = \boldsymbol{A}\boldsymbol{R}_\mathrm{z}\boldsymbol{A}^{\mathrm{H}}, \tag{5.24}$$

where $\boldsymbol{R}_\mathrm{z} = \int_{f\in\mathcal{F}_\mathrm{s}} \boldsymbol{z}(f)\boldsymbol{z}^{\mathrm{H}}(f)\mathsf{d}f$ is an $L\times L$ covariance matrix of $\boldsymbol{z}(f)$ that has at most $2K$ non-zero rows and columns. Following (2.40), we can now construct a frame $\boldsymbol{V}$ whose column space is the span of $\boldsymbol{y}(f)$ by decomposing $\boldsymbol{R}$ such that

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}. \tag{5.25}$$

Obtained frame $\boldsymbol{V}$ can then be used to find the support[6] $\mathcal{S}(\boldsymbol{z}(\mathcal{F}_\mathrm{s})) = \mathcal{S}(\boldsymbol{z}[n])$ by solving $\boldsymbol{V} = \boldsymbol{A}\boldsymbol{U}$ for the sparsest $\boldsymbol{U}$ [137], e.g., applying available MMV recovery algorithms such as the SOMP [134]. Once $\mathcal{S}(\boldsymbol{z}[n])$ is found, the elements of $\boldsymbol{z}[n]$ are given by

$$z_i[n] = \begin{cases} \boldsymbol{a}_i^{\mathrm{H}}\boldsymbol{a}_i\boldsymbol{y}[n], & i\in\mathcal{S}(\boldsymbol{z}[n]) \\ 0 & \text{otherwise} \end{cases}, \tag{5.26}$$

where $\boldsymbol{a}_n$ denotes the nth column of $\boldsymbol{A}$. Vector entries $z_i(f)$ can be found by computing the DTFT of $z_i[n]$. Alternatively, one can recover them directly from the DTFT of $\boldsymbol{y}[n]$.

Recovery with $f_\mathrm{s} = p(2k+1)f_\mathrm{p}$. Consider now the case when $f_\mathrm{s} = (2k+1)f_\mathrm{p}$ and $f_\mathrm{p}\approx B$ which corresponds to the right-hand side of Figure 5.8. We see that $\mathcal{F}_\mathrm{s}$ in this case is comprised of $p = 2k+1$ spectral cells of width f_d, each contacting different shifted copies of $X_k(f)$. Hence, we cannot directly represent $\boldsymbol{y}(f)$ via $\boldsymbol{z}(f)$ anymore. We note, however, that

$$y_m(f+if_\mathrm{p}) = \sum_{\ell=-L_0}^{L_0} c_{m,\ell}S(f-(\ell-k)f_\mathrm{p}) = \sum_{\ell=-L_0-i}^{L_0-i} c_{m,\ell+i}S(f-\ell f_\mathrm{p}), \tag{5.27}$$

where $i = -k,\ldots,0,\ldots,k$. This means that from the output $y_m[n]$ of a single physical

[6] We denote the overall support of $\boldsymbol{z}(f)$ by $\mathcal{S}(\boldsymbol{z}(\mathcal{F}_\mathrm{s}))$, whereas the notation $\mathcal{S}(\boldsymbol{z}(f))$ indicates the support with respect to each particular frequency $f\in\mathcal{F}_\mathrm{s}$.

sampling channel we can obtain $p \leq L$ virtual channel outputs $y_{m,i}$. We can do so by modulating $y_m[n]$ to if_{p}, low-pass filtering it with a cut-off $f_{\mathrm{p}}/2$ and then decimating the result by a factor of p.

Since all $y_{m,i}[n]$ result from a single mixing sequence $p_m(t)$, this yields a partially circulant structure of the sensing matrix $\boldsymbol{A}$ whose $(m-k)$th to $(m+k)$th rows are given by

$$[\boldsymbol{A}]_{m-k:m+k} = \begin{pmatrix} c_{m,L_0-k+1} & \cdots & c^*_{m,L_0} & \cdots & c_{m,-k} & \cdots & c_{m,L_0-k} \\ c^*_{m,L_0} & \cdots & c^*_{m,L_0-k} & \cdots & c_{m,0} & \cdots & c_{m,L_0} \\ c^*_{m,L_0-k} & \cdots & c^*_{m,L_0-2k} & \cdots & c_{m,k} & \cdots & c^*_{m,L_0-k+1} \end{pmatrix}. \tag{5.28}$$

Once $y_{m,i}$ are obtained, we can form (5.22) by collecting them together into a single vector

$$\boldsymbol{y}[n] = [y_{1,-k}[n], y_{1,0}[n], y_{1,k}[n], y_{2,-k}[n], \ldots, y_{M,k}[n]]^{\mathrm{H}}$$

of length $M_{\mathrm{t}} = pM$, where M is the number of physical sampling channels according to the block-scheme from Figure 5.6, and composing $\boldsymbol{A}$ from $c_{m,\ell}$ according to (5.28). From this point on, the recovery of $\boldsymbol{z}(f)$ from $\boldsymbol{y}[n]$ can proceed as before in the basic configuration.

5.3.1.5 PSD recovery

It has been recently shown that besides reconstructing the full spectrum $S(f)$ (or equivalently the time-domain signal $s(t)$), one could also directly recover the signal's PSD from the sub-Nyquist samples $\boldsymbol{y}[n]$, provided that $s(t)$ is wide-sense stationary [256], [257], [282]. Note that the power spectrum of $s(t)$ is the Fourier transform of its autocorrelation function

$$P_s(f) = \int_{-\infty}^{\infty} r_s(\tau) e^{-\jmath 2\pi f \tau} \mathsf{d}\tau = \int_{-\infty}^{\infty} \mathbb{E}\left\{s(t)s^*(t-\tau)\right\} e^{-\jmath 2\pi f \tau} \mathsf{d}\tau = \mathbb{E}\left\{|S(f)|^2\right\}. \tag{5.29}$$

Consequently, the frequency supports of $P_s(f)$ and $S(f)$ are the same. We also observe that

$$\mathbb{E}\left\{S(f_1)S^*(f_2)\right\} = \begin{cases} P_s(f_1), & f_1 = f_2 \\ 0, & \text{otherwise} \end{cases}. \tag{5.30}$$

Consider now the spectral autocorrelation matrix $\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)$ defined as

$$\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f) = \mathbb{E}\left\{\boldsymbol{z}(f)\boldsymbol{z}^{\mathrm{H}}(f)\right\}, \tag{5.31}$$

where the elements of $\boldsymbol{z}(f)$ are given by (5.20). Similar to the covariance matrix $\boldsymbol{R}_{\mathrm{z}}$ in (5.24), $\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)$ is a diagonal $L \times L$ matrix. However, while the ℓth diagonal element of $\boldsymbol{R}_{\mathrm{z}}$ contains the energy of the ℓth spectral cell, the diagonal elements of $\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)$ are given by

$$[\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}]_{(\ell,\ell)}(f) = P_s(f - \ell f_{\mathrm{p}}). \tag{5.32}$$

Given (5.32), we can relate the unknown PSD $P_s(f)$ to the known correlation functions of the compressed measurements $\boldsymbol{y}[n]$ as

$$\boldsymbol{R}^{(\mathrm{a})}(f) = \mathbb{E}\left\{\boldsymbol{y}(f)\boldsymbol{y}^{\mathrm{H}}(f)\right\} = \sum_{n_\tau=-\infty}^{\infty} \boldsymbol{R}_y[n_\tau]e^{-\jmath 2\pi f n_\tau} = \boldsymbol{A}\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)\boldsymbol{A}^{\mathrm{H}}, \tag{5.33}$$

where $\boldsymbol{R}_y[n_\tau] = \mathbb{E}\left\{\boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n-n_\tau]\right\}$ and n_τ represents the (discrete) delay.

Due to the fact that $\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)$ is diagonal, we can vectorize $\boldsymbol{R}^{(\mathrm{a})}(f)$ to obtain

$$\boldsymbol{r}^{(\mathrm{a})}(f) = \mathrm{vec}\left(\boldsymbol{R}^{(\mathrm{a})}(f)\right) = (\boldsymbol{A}^* \otimes \boldsymbol{A})\boldsymbol{r}_{\mathrm{z}}^{(\mathrm{a})}(f) = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{r}_s(f), \tag{5.34}$$

where $\boldsymbol{r}_{\mathrm{z}}^{(\mathrm{a})}(f) = \mathrm{vec}\left(\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)\right)$, $\boldsymbol{r}_s(f) = \mathrm{diag}\left(\boldsymbol{R}_{\mathrm{z}}^{(\mathrm{a})}(f)\right)$, while $\otimes$ and $\odot$ denote the Kronecker and the Khatri-Rao products, respectively. What is left is to determine $\boldsymbol{r}_s(f)$ from $\boldsymbol{r}^{(\mathrm{a})}(f)$ for each $f \in \mathcal{F}_{\mathrm{p}}$. We notice that (5.34) is similar to (5.19), since each $\boldsymbol{r}_s(f)$ is at most K-sparse. We can approach it the same way we have approached (5.19), e.g., by applying the CTF block (5.23)-(5.25). Furthermore, it has also been noted that, if $N/2 < M$, (5.34) is over-determined and, as such, can be solved without the sparsity constraint [256], [257].

5.3.1.6 Sensing matrix analysis

Written in a matrix form, the relation between the (discrete) low-rate output of the MWC and its wideband analog input can be seen as a rather typical CS system. To solve it, we can either consider each output sample $\boldsymbol{y}[n]$ individually or apply the MMV recovery strategy that employs the CTF block, as described above. In any case, the sensing matrix $\boldsymbol{A}$ in (5.22) has a similar meaning as in a regular CS system considered in the first part of this thesis. With traditional sensing matrix CS metrics, such as the RIP and the matrix coherence, being examined in [288], here we briefly look into the spread of the OSNR in the MWC.

According to the analysis performed in Chapter 3, the application of $\boldsymbol{A}$ to $\boldsymbol{z}(f)$ results in a spread of the effective signal energy in $\boldsymbol{y}(f)$ with respect to the support $\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))$. Formally, we define the signal energy at the output of the MWC as

$$E_{\mathrm{O}} = \left\| \int\limits_{f\in\mathcal{F}_{\mathrm{p}}} |\boldsymbol{A}\boldsymbol{z}(f)|^2 \mathsf{d}f \right\|_1, \tag{5.35}$$

whereas the input signal energy is given by

$$E_{\mathrm{I}} = \int\limits_{f\in\mathcal{F}} |S(f)|^2 \mathsf{d}f = \left\| \int\limits_{f\in\mathcal{F}_{\mathrm{p}}} |\boldsymbol{z}(f)|^2 \mathsf{d}f \right\|_1 = \sum_{\ell} E_\ell, \tag{5.36}$$

where $E_\ell = \int\limits_{f\in\mathcal{F}_{\mathrm{p}}} |z_\ell(f)|^2 \mathsf{d}f$.

Single active sub-band. Consider a single active sub-band with a central frequency f_c, a bandwidth B and a total energy $E_I = \int_{f\in\mathcal{F}_1} |X_1(f)|^2 \mathsf{d}f$. When $f_c = qf_p$, $X_1(f)$ contributes to a single non-zero entry of $\boldsymbol{z}(f)$ only, namely to $z_q(f)$. Then, (5.35) transforms into

$$E_O = \sum_{m=1}^{M} \int_{f\in\mathcal{F}_p} |a_{m,q} z_q(f)|^2 \mathsf{d}f = \sum_{m=1}^{M} |a_{m,q}|^2 \int_{f\in\mathcal{F}_p} |z_q(f)|^2 \mathsf{d}f = \|\boldsymbol{a}_q\|_2^2 E_I, \tag{5.37}$$

where $\boldsymbol{a}_q$ indicates the qth column of $\boldsymbol{A}$. We notice that, as to be expected, (5.37) corresponds to the result (3.15) obtained for the case of a 1-sparse input in a canonical CS setting. When $f_c \neq qf_p$ the signal bandwidth B gets split between two adjacent cells with indices q_1, q_2 in a ratio B_1/B_2 where $B_1 + B_2 = B$. For a signal with a rectangular energy spectrum envelope, this means that $E_{q_1} = \frac{B_1}{B} E_I = E_I - E_{q_2}$. For an energy spectrum with an arbitrary envelope $g(f), f \in \mathcal{F}_p$, we generally have that $E_{q_1} = \frac{\int_{-B/2}^{-B/2+B_1} |g(f)|^2 \mathsf{d}f}{\int_{-B/2+B_1}^{B/2} |g(f)|^2 \mathsf{d}f} E_{q_2}$.

Since $f_p \geq B$, the two signal parts do not overlap in $\mathcal{F}_p$ which allows us to compute E_O

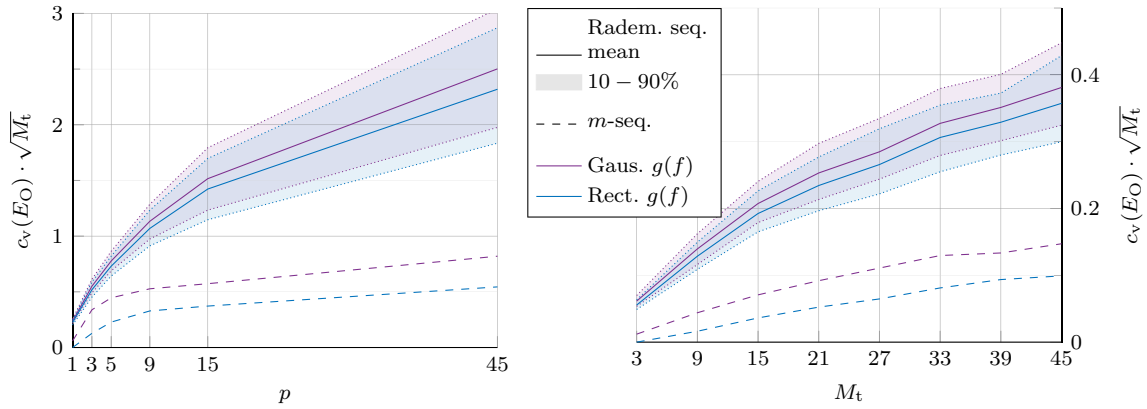
$$\begin{aligned} E_O &= \sum_{m=1}^{M} \int_{f\in\mathcal{F}_p} |a_{m,q_1} z_{q_1}(f) + a_{m,q_2} z_{q_2}(f)|^2 \mathsf{d}f = \sum_{m=1}^{M} \int_{f\in\mathcal{F}_p} \left(|a_{m,q_1} z_{q_1}(f)|^2 + |a_{m,q_2} z_{q_2}(f)|^2\right) \mathsf{d}f \\ &= \sum_{m=1}^{M} \left(|a_{m,q_1}|^2 E_{q_1} + |a_{m,q_2}|^2 E_{q_2}\right) = \|\boldsymbol{a}_{q_1}\|_2^2 E_{q_1} + \|\boldsymbol{a}_{q_2}\|_2^2 E_{q_2}. \end{aligned} \tag{5.38}$$

Assuming rectangular spectral mask, we simply have that $E_o = \left(\||\boldsymbol{a}_{q_1}|\|_2^2 \frac{B_1}{B} + \||\boldsymbol{a}_{q_2}|\|_2^2 \frac{B_2}{B}\right) E_I$.

Expressions (5.37), (5.38) show that in the case of a single active sub-band the output signal energy depends on the support of $\boldsymbol{z}(f)$ through the energies of the corresponding Fourier coefficients of $p_m(t)$ (note that in the basic setting $|a_{m,q}|^2 = |c_{m,q}|^2$). Furthermore, in the case of sub-band splitting, for a fixed support the actual value of E_O depends on the proportion in which the signal energy gets split between the spectral cells.

Example 5.20

In this example, we numerically investigate the spread of the output signal energy in the MWC for the case of $K = 1$ active signal sub-bands with $f_p \approx B$. We do so for two choices of mixing sequences $p_m(t)$: pseudo-random Rademacher sequences (± 1 sequences with equal probability of observing $+1$ or -1) and maximum-length or simply m-sequences obtained as an output of a linear feedback n-bit shift register with an exclusive-or (XOR) gate on the tap bits [289]. In both cases, the sequence length is set to $L = 2^n - 1 = 127$ while the total number of output channels is $M_t = pM$. For the m-sequence generation, to obtain M different sequences we use M different (random) register initializations. We then evaluate the energy spread for 100 realizations of the sequence sets $\{p_m(t)\}_{m=1}^{M}$.



(a) $c_\mathrm{v}(E_\mathrm{O})$ vs. p for $M_\mathrm{t} = 45$

(b) $c_\mathrm{v}(E_\mathrm{O})$ vs. M_t for $M = 3$

Figure 5.9: Spread of the signal energy at the output of the MWC depending on the choice of the mixing sequences $p_m(t)$ for $L = 127$ and $K = 1$: $c_\mathrm{v}(E_\mathrm{O})$ as a function of the expanding coefficient p for $M_\mathrm{t} = pM = 45$ (a) and $c_\mathrm{v}(E_\mathrm{O})$ as a function of M_t for $M = 3$. The value of the expanding coefficient p defines the number of physical sampling channels used, e.g., $p = 1$ means that $M = M_\mathrm{t} = 45$, while for $p = 45$ and $M_\mathrm{t} = 45$ we have that $M = 1$.

Figure 5.9a demonstrates the mean and the $(10 - 90)\%$ area of the coefficient of variation of E_O over the central frequency f_c of the active sub-band as a function of the expanding coefficient p for a fixed total number of channels $M_\mathrm{t} = pM = 45$. Figure 5.9b, on the other hand, shows $c_\mathrm{v}\left(E_\mathrm{O}\right)$ as a function of $M_\mathrm{t} = pM$ for a fixed number of physical sampling channels $M = 3$. Note that $p = 1$ means that $f_\mathrm{s} = f_\mathrm{p}$ and $M = M_\mathrm{t}$, whereas $p = M_\mathrm{t}$ implies that a single physical sampling channel is expanded into $M_\mathrm{t} = p$ virtual ones. For comparison, we show the results for two different spectral masks $g(f)$, a Guassian and a rectangular one. First, we see that different sets of Rademacher mixing sequences vary significantly with respect to the spread of the signal's energy, which indicates that the choice of an actual realization makes a difference when one wishes to minimize the effect of the OSNR spread in the MWC. For m-sequences, however, the situation is radically different. Since m-sequences are designed to have a flat spectrum, for any given n all Fourier coefficients $c_{m,\ell}$ have the same magnitude but for $c_{m,0}$, which is $2^{\frac{n}{2}}$ times smaller than the rest. Therefore, it is irrelevant which particular realization is taken: they all have exactly the same $|c_{m,\ell}|$. Furthermore, this means that in case of a single active sub-band the spread of the signal's energy with m-sequences occurs only due to this single direct current (DC) component $c_{m,0}$. In other words, only a small fraction of possible input signals is subject to an attenuation at the output, while for the majority of possible inputs $E_\mathrm{O} = E_\mathrm{I}$, which shows itself in lower values of $c_\mathrm{v}(E_\mathrm{O})$. Interestingly, for both types of mixing sequences the energy spread generally increases for larger values of the expanding coefficient p. Finally, we see that the spread of E_O depends not only on the realization of the mixing and sampling steps but also on the type of the spectral mask used.

Multiple active sub-bands. In case of multiple active sub-bands, we have that

$$E_{\mathrm{O}} = \sum_{m=1}^{M} \int_{f\in\mathcal{F}_{\mathrm{p}}} \left| \sum_{q\in\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))} a_{m,q} z_q(f) \right|^2 \mathsf{d}f, \tag{5.39}$$

which in general case one cannot simplify any further. We could, however, upper bound E_{O}

$$\begin{aligned} E_{\mathrm{O}} &\leq \sum_{m=1}^{M} \int_{f\in\mathcal{F}_{\mathrm{p}}} \left(\sum_{q\in\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))} |a_{m,q} z_q(f)| \right)^2 \mathsf{d}f \leq \sum_{m=1}^{M} \int_{f\in\mathcal{F}_{\mathrm{p}}} \sum_{q\in\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))} |a_{m,q} z_q(f)|^2 \mathsf{d}f \\ &= \sum_{q\in\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))} \sum_{m=1}^{M} |a_{m,q}|^2 \int_{f\in\mathcal{F}_{\mathrm{p}}} |z_q(f)|^2 \mathsf{d}f = \sum_{q\in\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))} \|\boldsymbol{a}_q\|_2^2 E_q. \end{aligned} \tag{5.40}$$

5.3.2 Direct compressive coarse energy sensing

Expression (5.19) binds the known spectrum of the low-rate MWC output to the unknown spectra of the bandpass signal components $z_\ell(f) = S(f-\ell f_{\mathrm{p}}), f \in \mathcal{F}_{\mathrm{p}}$. Therefore, determining the support of $z_\ell(f)$ on $\mathcal{F}_{\mathrm{p}}$ is tantamount to coarse spectrum sensing at the resolution of f_{p} [276]. Recall the discussion of the coarse WSS in Section 5.2.2.1. Representing $S(f)$ via $\boldsymbol{z}(f)$ is equivalent to dividing W into a set of sensing channels $\mathcal{C}_n$ (e.g., as in Figure 5.3) with bandwidths f_{p} and central frequencies ℓf_{p}. The difference is that the sensing resolution in the MWC is defined by the hardware setting of the mixing sequences' period T_{p} rather than by applying a (digital) filter-bank whose parameters can be adjusted arbitrarily. Thus, one could use the spectrum recovery strategy described above to determine the support of $\boldsymbol{z}(f)$, that, in terms of the coarse WSS, determines the indices of the occupied sensing channels $\mathcal{C}_n$. One could also first reconstruct the full spectrum $S(f)$ or the PSD $P(f)$ and then perform the sensing, e.g., coarse or fine [257], [274], [275], [280], [290]. Here, we propose an alternative approach where we formulate the recovery step directly as a coarse energy detection problem. It avoids full signal/spectral recovery and, similarly to the case of PSD recovery, can potentially by applied with a densely occupied spectrum.

From (5.24), it is easy to see that the diagonal elements of $\boldsymbol{R}_{\mathrm{z}} = \int_{f\in\mathcal{F}_{\mathrm{p}}} \boldsymbol{z}(f)\boldsymbol{z}^{\mathrm{H}}(f)\mathsf{d}f$ contain the energies of the corresponding bandpass signal components since

$$[\boldsymbol{R}_{\mathrm{z}}]_{(\ell,\ell)} = \int_{f\in\mathcal{F}_{\mathrm{p}}} |z_\ell(f)|^2 \mathsf{d}f = \int_{f\in\mathcal{F}_{\mathrm{p}}} |S(f-\ell f_{\mathrm{p}})|^2 \mathsf{d}f = E_\ell. \tag{5.41}$$

The non-diagonal elements $[\boldsymbol{R}_{\mathrm{z}}]_{k,j:k\neq j}$, on the other hand, represent cross-correlation terms between kth and jth signal components. Hence, when the signals within the individual sub-bands are uncorrelated, $\boldsymbol{R}_{\mathrm{z}}$ becomes a diagonal matrix with $\tilde{K} \leq 2K$ non-zero elements.

Vectorizing the covariance matrix $\boldsymbol{R} = \sum_{n=-\infty}^{\infty} \boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n]$, we can obtain

$$\boldsymbol{r} = \mathrm{vec}(\boldsymbol{R}) = (\boldsymbol{A}^* \otimes \boldsymbol{A})\boldsymbol{r}_{\mathrm{z}} = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{\varepsilon}, \tag{5.42}$$

where $\boldsymbol{r}_{\mathrm{z}} = \mathrm{vec}(\boldsymbol{R}_{\mathrm{z}})$ and $\boldsymbol{\varepsilon} = \mathrm{diag}(\boldsymbol{R}_{\mathrm{z}})$ is a $\tilde{K}$-sparse vector of length L that contains the energies E_ℓ, i.e., $\varepsilon_\ell = \int_{f\in\mathcal{F}_{\mathrm{p}}} |z_\ell(f)|^2 \mathsf{d}f$ where $\ell = -L_0, \ldots, L_0$. We notice that (5.42) is reminiscent of (5.34). However, contrary to (5.34), here it is a finite-dimensional problem that can be solved at once, e.g., we can recover $\boldsymbol{\varepsilon}$ by solving the following recovery problem

$$\hat{\boldsymbol{\varepsilon}} = \arg\left(\min_{\boldsymbol{\varepsilon}} \|\boldsymbol{\varepsilon}\|_1\right) \text{ s.t. } \boldsymbol{r} = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{\varepsilon}. \tag{5.43}$$

where $\boldsymbol{r} = \mathrm{vec}(\boldsymbol{R})$ and $\boldsymbol{\varepsilon} = \mathrm{diag}(\boldsymbol{R}_{\mathrm{z}})$.

Consider now an $L \times 1$ binary occupancy vector $\boldsymbol{b}$ from (5.6) with elements that are now defined as

$$b_\ell = \begin{cases} 1, & \text{if } E_\ell > 0 \\ 0, & \text{otherwise} \end{cases}. \tag{5.44}$$

We have that $b_\ell = 0\ \forall \ell : E_\ell = 0$ and $\|\boldsymbol{b}\|_0 = |\mathcal{S}(\boldsymbol{\varepsilon})|$ where $\mathcal{S}(\boldsymbol{\varepsilon})$ contains the indices of the spectral cells that have signal energy. Therefore, once $\mathcal{S}(\boldsymbol{\varepsilon})$ is estimated one can detect the presence of the signal's energy in the ℓth sensing channel $\mathcal{C}_\ell$ (represented by the ℓth spectral cell), either by directly interpreting $\mathcal{S}(\boldsymbol{\varepsilon})$ as the detector output or by additionally comparing the recovered energies $\hat{\boldsymbol{\varepsilon}}_{\mathcal{S}(\boldsymbol{\varepsilon})}$ with a predefined threshold. This yields two coarse multiband energy detectors (MEDs), decision rules for which are given below.

Non-parametric direct decision rule (DDR). Since the support $\mathcal{S}(\hat{\boldsymbol{\varepsilon}})$ contains the positions of the non-zero entries of $\hat{\boldsymbol{\varepsilon}}$, the decision rule for signal detection can be simply given by

$$\hat{b}_\ell = \begin{cases} 1, & \text{if } \ell \in \mathcal{S}(\hat{\boldsymbol{\varepsilon}}) \\ 0, & \text{otherwise} \end{cases}. \tag{5.45}$$

Note that (5.45) yields a non-parametric detector as the binary occupancy vector $\boldsymbol{b}$ is obtained as a by-product of the recovery step (5.43). In the following, we refer to (5.45) as the C-DDR where C stands for compressive.

An application of C-DDR results in the coarse estimation of the spectral support

$$\mathcal{F}_{\mathrm{s}}^{\mathrm{c}} = \bigcup_{i\in\mathcal{S}(\hat{\boldsymbol{\varepsilon}})} \left[\frac{2i-1}{2} f_{\mathrm{p}}, \frac{2i+1}{2} f_{\mathrm{p}}\right].$$

Similarly to (5.5), when the support $\mathcal{S}(\hat{\boldsymbol{\varepsilon}})$ is estimated correctly (i.e., $\mathcal{S}(\hat{\boldsymbol{\varepsilon}}) = \mathcal{S}(\boldsymbol{\varepsilon})$), the difference δ_f between the actual and the course spectral supports is bounded as

$$0 \leq \delta_f = \lambda^*(\mathcal{F}_{\mathrm{s}}^{\mathrm{c}}) - \lambda^*(\mathcal{F}_{\mathrm{s}}) \leq K f_{\mathrm{p}}. \tag{5.46}$$

In the presence of erroneous estimates when $\mathcal{S}(\hat{\boldsymbol{\varepsilon}}) \neq \mathcal{S}(\boldsymbol{\varepsilon})$, a false alarm occurs in case of $\mathcal{S}(\hat{\boldsymbol{\varepsilon}}) \backslash \mathcal{S}(\boldsymbol{\varepsilon}) \neq \emptyset$ whereas a missed detection happens when $\mathcal{S}(\boldsymbol{\varepsilon}) \backslash \mathcal{S}(\hat{\boldsymbol{\varepsilon}}) \neq \emptyset$. Hence, the spectral occupancy estimation error δ_f in this case is

$$0 < \delta_f \leq \left(|\mathcal{S}(\boldsymbol{\varepsilon}) \ominus \mathcal{S}(\hat{\boldsymbol{\varepsilon}})| + |\mathcal{S}(\boldsymbol{\varepsilon}) \cap \mathcal{S}(\hat{\boldsymbol{\varepsilon}})| \right) f_{\mathrm{p}}. \tag{5.47}$$

Energy-based decision rule (EDR). Given the vector of energy estimates $\hat{\boldsymbol{\varepsilon}}$, we can also apply a classical thresholding, such as in (5.3), to determine the entries of the binary occupancy vector $\boldsymbol{b}$, namely

$$\hat{b}_\ell = \begin{cases} 1, & \text{if } \hat{\varepsilon}_\ell > \zeta \\ 0, & \text{otherwise} \end{cases}, \tag{5.48}$$

where ζ is some threshold whose value is defined in advance. Note that (5.48) converges to (5.45) for $\zeta \to 0$. Following our notations, we refer to (5.48) as C-EDR.

The main difference of C-EDR from C-direct decision rule (DDR) is that, although a false alarm in C-EDR can still only occur when there are erroneous support estimate entries, a missed detection is now possible even when the support is estimated correctly. On the other hand, when there are erroneous entries in $\mathcal{S}(\hat{\boldsymbol{\varepsilon}})$ the thresholding in (5.48) allows reducing the false alarm probability below what is achievable by applying (5.45).

5.3.3 Numerical study

To evaluate the performance of the coarse wideband sensing from sub-Nyquist samples, we perform a series of numerical experiments comparing different MEDs.

5.3.3.1 Simulation setup

We generate test multiband signals $S(f)$ in the frequency domain by representing the Nyquist range W with a dense grid of LN equidistant points. Each test signal is comprised of K non-overlapping active sub-bands of equal bandwidth B such that

$$S[nf_{\mathrm{g}}] = \begin{cases} \sigma_{\mathrm{s},k} g[nf_{\mathrm{g}} - f_{\mathrm{c},k}] \mathrm{e}^{\jmath\phi[nf_{\mathrm{g}}]}, & n \in [a_k, b_k] \\ 0, & \text{otherwise} \end{cases}, \tag{5.49}$$

where $f_{\mathrm{g}} = \frac{W}{LN}$ is the frequency grid step, $g[nf_{\mathrm{g}}]$ represents a (discrete) spectral mask, and $a_k = \left\langle \frac{f_{\mathrm{c},k} - B/2}{f_{\mathrm{g}}} \right\rangle$, $b_k = \left\langle \frac{f_{\mathrm{c},k} + B/2}{f_{\mathrm{g}}} \right\rangle$, $n \in \mathbb{Z}$. The values of ϕ, $\sigma_{\mathrm{s},k}$ and $f_{\mathrm{c},k}$ are drawn independently at random from uniform distributions $\mathcal{U}(-\pi, \pi)$, $\mathcal{U}(0.7, 1)$ and $\mathcal{U}(B/2, W - B/2)$, respectively. The elements of the $M \times L$ sensing matrix $\boldsymbol{A}$ are computed as Fourier coefficients of the underlying mixing sequences that are generated as i.i.d. Rademacher random variables. The total number of sampling channels is $M = 30$, which yields the total sampling rate that is $30/127 \approx 25\%$ of Nyquist. In the presence of wideband additive

Signal parameters	**Sampling parameters**	**Considered multiband sensing methods**	
$K \in [1, 6]$	$L = 127 \leq W/B$	C-DDR (5.45)	proposed compressive MED
$W/B = 132$	$f_{\mathrm{p}} = \frac{W}{L}$	C-EDR (5.48)	proposed compressive MED
$g[nf_{\mathrm{g}}] = e^{-\frac{18(nf_{\mathrm{g}})^2}{B^2}}$	$f_{\mathrm{s}} = f_{\mathrm{p}}$	C-PSD [257]	compressive PSD-based MED
$N = 51$	$M = 30$	Nyq-ref	Nyquist-rate MED

Table 5.2: List of simulation parameters and considered multiband sensing methods

Total observed bandwidth	W	GHz	2	2	5	5
Sub-band bandwidth	$B \geq B_k$	MHz	15.15	20	20	10
Number of spectral cells	$L \leq W/B$		127	95	247	495
Sampling rate	$f_{\mathrm{s}} = f_{\mathrm{p}} = W/L$	MHz	15.75	21.05	20.25	10.1
Total sampling rate	Mf_{s}	MHz	472.5	631.5	607.29	303
Mixing sequence bandwidth	Lf_{p}	GHz	2	2	5	5

Table 5.3: Examples of possible parameter values for different choices of W and L.

noise $w[nf_{\mathrm{g}}]$, we model $w[nf_{\mathrm{g}}]$ as i.i.d. circularly symmetric complex Gaussian variables with variance σ_0^2 where the input SNR is defined as $\eta_{\mathrm{I}} = E_{\mathrm{I}}/KN\sigma_0^2$. The values of the parameters used throughout the simulations can be found in Table 5.2. Note that for performance evaluation of multiband detection the actual values of W and B are largely irrelevant. Nevertheless, in Table 5.3 we give some examples of possible parameter values that result from different choices of W and L.

We study the performance of the proposed compressive energy detectors C-DDR and C-EDR by comparing them with related coarse wideband sensing methods. One of these is an energy-based multiband coarse sensing performed on the signal's PSD reconstructed from the sub-Nyquist samples $\boldsymbol{y}[n]$ [257], which we refer to as C-PSD. In C-PSD, we first reconstruct the PSD of the input signal according to the procedure presented in Section 5.3.1.5. We then split the recovered PSD into L non-overlapping consecutive blocks of bandwidth f_{p} and apply the energy-based sensing in each of them, as described in Section 5.2.2.1. For PSD recovery from sub-Nyquist samples in C-PSD we apply the SOMP, whereas for energy recovery in C-DDR and C-EDR we use the OMP. The number of the active signal sub-bands is assumed to be known and is used for defining a stopping criterion[7] for the recovery algorithms. Note that we do not specifically consider here an approach where we first reconstruct the time-domain Nyquist-rate signal samples and then compute the frequency or power spectrum from them, since it has been already shown to provide a worse performance than direct

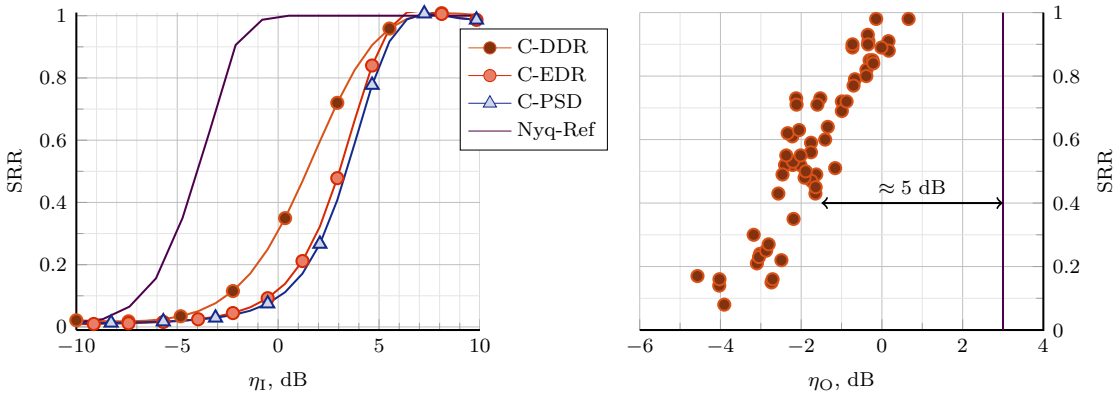
[7]Note that since the sparsity order is upper bounded by $2K$ we use double the number of sub-bands as a maximum number of iterations.

PSD recovery as in C-PSD [257]. As a second reference method, we consider energy-based multiband sensing from the Nyquist-rate input, which proceeds according to the same coarse detection strategy as in C-PSD; we refer to it as Nyq-Ref. The four considered MEDs are also summarized in Table 5.2.

Finally, we note that the application of C-EDR, C-PSD and Nyq-Ref requires setting a detection threshold ζ. Conventionally, the threshold for energy detection is chosen based on a desired false alarm rate (in case of wideband multiband sensing that would be per channel false alarm rate), provided that the noise statistics is known at the detector. In a sub-Nyquist multiband setting, however, such an approach encounters certain difficulties. Thus, while the input noise statistics might be available, at the detection step one is dealing with a reconstruction error rather than the input noise itself. Therefore, instead of attempting to evaluate the "noise" PDF after the reconstruction, we rely on the assumption that the power of the reconstruction error is bounded [18] in proportion to the input noise level and set the threshold ζ as $\zeta = \kappa N \sigma_0^2$ where $\kappa > 0$ is varied.

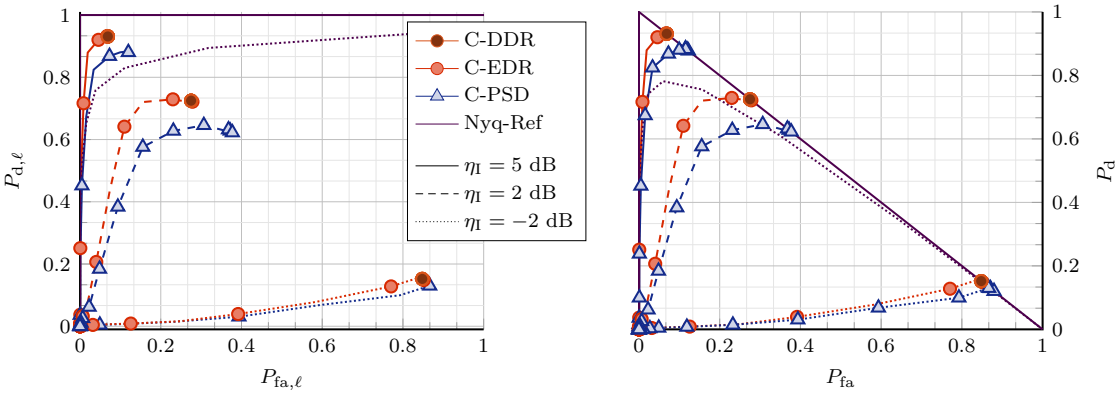
5.3.3.2 Single occupied spectral cell

We begin by considering a single active signal sub-band with a central frequency that is an odd multiple of $f_{\mathrm{p}}/2$, meaning that it always occupies exactly one spectral cell, i.e., $|\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))| = K = 1$. We first study the SRR defined by (5.12). Figure 5.10a shows average SRR as a function of the ISNR for Rademacher mixing sequences. We see that the Nyq-Ref detector is expectedly superior to compressive methods. Among the compressive detectors on the other hand, the C-DDR outperforms the rest while C-EDR and C-PSD demonstrate largely similar performance. Note however, that these and all the following results for Nyq-Ref and both compressive methods using thresholding correspond to the case when the noise variance is perfectly known to the detector, which provides a possibility to set the detection threshold appropriately. In more practical scenarios of noise uncertainty, the performance of the threshold-based detectors is likely to deteriorate, which is not the case for C-DDR as it does not explicitly rely on the knowledge of the noise statistics. On the other hand, for compressive approaches we expect performance variations over the support of the input signal (the central frequencies of the signal sub-bands) due to the spread of the effective SNR at the receiver output, as discussed in Chapter 3. Thus, Figure 5.10b depicts the SRR for the C-DDR detector against the OSNR in a form of a scatter plot at a fixed η_{I} value, e.g., each point on the graph represents a distinct signal support with a corresponding η_{O}. One can clearly discern a familiar trend here, namely that higher effective SNRs at the receiver output tend to yield higher recovery rates. Note that for Nyq-Ref we have that $\eta_{\mathrm{O}} = \eta_{\mathrm{I}} = 3$ dB while SRR $= 1$ irrespective of the signal support. Interestingly, the average effective SNR difference between the Nyquist and sub-Nyquist rate systems observed in Figure 5.10b (≈ 5 dB) corresponds to an average difference in the ISNR between C-DDR and Nyq-ref SRR curves from Figure 5.10a.



(a) average SRR vs. η_{I} **(b)** support-specific SRR vs. η_{O}

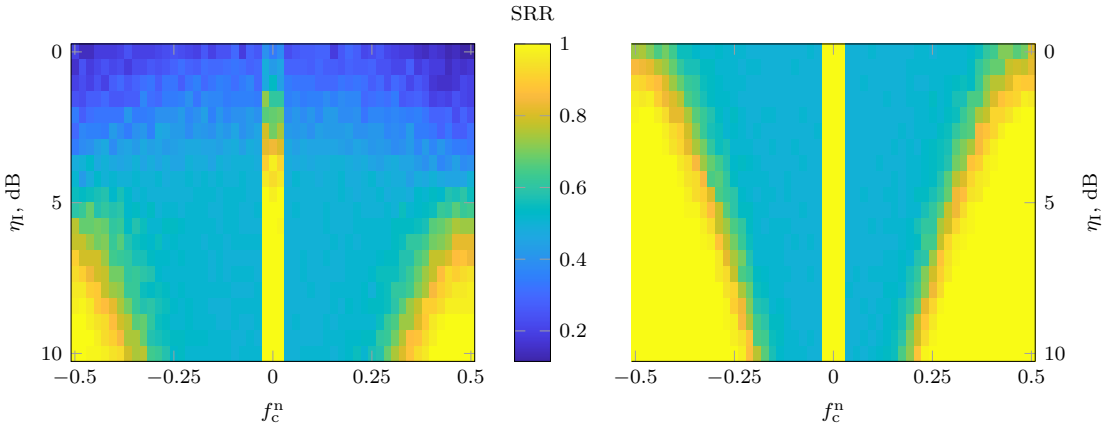
Figure 5.10: Support recovery rate as a function of the SNR for Rademacher mixing sequences with $p = 1$ and a single active sub-band occupying a single spectral cell: average SRR vs. ISNR (a) and support-specific SRR for C-DDR vs. corresponding OSNR for $\eta_{\mathrm{I}} = 3$ dB. For detection threshold we set $\kappa = 0.1$.



(a) individual ROC **(b)** total ROC

Figure 5.11: Individual (per sensing channel) (a) and total (b) ROC curves for Rademacher mixing sequences with $p = 1$ and several ISNR values.

We now turn to examining the false alarm and correct detection probabilities. Figure 5.11a presents the average (among sensing channels) receiver operating characteristicss (ROCs), i.e., a posterior probability of correct detection $P_{\mathrm{d},\ell}$ versus a posterior probability of false alarm $P_{\mathrm{fa},\ell}$, per sensing channel for three different values of ISNR. Note that since C-DDR does not require setting a threshold, it is represented by a single point for each SNR. We also note that probabilities of false alarm and correct detection for C-EDR, C-PSD do not span the whole range from zero to one. The reason behind this is that their performance is bound to the support recovery performance of the underlying recovery method, i.e., a false alarm can only occur if $\mathcal{S}(\hat{\boldsymbol{z}}(\mathcal{F}_{\mathrm{p}}))\backslash\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}})) \neq \emptyset$. A missed detection, on the other hand, can happen either if $\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))\backslash\mathcal{S}(\hat{\boldsymbol{z}}(\mathcal{F}_{\mathrm{p}})) \neq \emptyset$ or if $\mathcal{S}(\boldsymbol{z}(\mathcal{F}_{\mathrm{p}}))\backslash\mathcal{S}(\hat{\boldsymbol{z}}(\mathcal{F}_{\mathrm{p}})) = \emptyset$ and the detection threshold ζ is set too high. This also explains why, at equal SNRs, both the false alarm and correct detection rates for C-EDR tend to decrease with the increase of the detection threshold from the value produced by C-DDR. Contrary, the reference Nyq-Ref detection



(a) C-DDR **(b)** Nyq-Ref

Figure 5.12: Support recovery rate vs. ISNR vs. relative position of the signal sub-band $f_{\mathrm{c}}^{\mathrm{n}} = f_{\mathrm{c}} - \left\lceil \frac{f_{\mathrm{c}}}{f_{\mathrm{p}}} \right\rfloor f_{\mathrm{p}}$ with respect to the center of the spectral cell nf_{p} for C-DDR (a) and Nyq-Ref (b) detectors.

scheme that corresponds to the Nyquist-rate sampling has rather traditional ROC curves that support its superiority to compressive methods. However, the ROC curves for the total posterior probabilities of correct detection P_{d} and false alarm P_{fa} presented in Figure 5.11b show already a different trend. For the reference Nyquist-sampled scheme it is characterized by the inversion of the characteristic for high false alarm rates and overall performance deterioration with regard to the SNR that is consistent with what we observed in our toy example in Figure 5.5. Contrarily, the character of ROC curves for C-DDR, C-EDR and C-PSD does not seem to be affected. This is largely due to the above mentioned property of sparse recovery based detectors which manifests itself in the fact that a detection error occurs only in the case of a wrongly estimated support.

5.3.3.3 Two occupied spectral cells.

So far, we have considered a single signal sub-band that occupies a single spectral cell, meaning that no energy splitting is taking place. In order to study how the position of the signal sub-band influences the detection performance, we change the central frequency of the active signal sub-band with respect to the center of the nearest spectral cell and compute the SRR for different values of the ISNR, which we depict in Figure 5.12 in a form of a color plot: Figure 5.12a shows the results for the proposed C-DDR detector, while Figure 5.12b presents SRR for Nyq-Ref scheme. In line with our expectations, the best performance with respect to SNR for both detectors corresponds to $f_{\mathrm{c}} \in \{nf_{\mathrm{p}}\}_{n=1}^{N}$, i.e. when the whole sub-band's energy is contained within exactly one spectral cell. However, it drops rapidly with the deviation from this optimal position, while starting to raise again after the relative central frequency exceeds a certain value, albeit at higher SNRs. To understand this peculiar behaviour, let us examine what happens when the signal sub-band center starts moving away from the center of the spectral cell (aka sensing channel) in greater detail. We saw that when $f_{\mathrm{p}} \approx B$ and

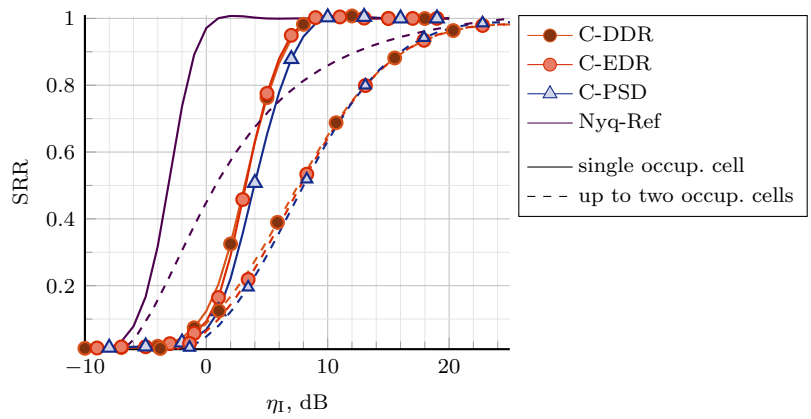


Figure 5.13: Support recovery rate for a single active sub-band that either occupies a single spectral cell or has an arbitrary central frequency chosen uniformly at random from $\mathcal{F}$.

$f_{\mathrm{c}} \notin \{nf_{\mathrm{p}}\}_{n=1}^{N}$ the signal energy gets split between two neighbouring spectral cells. This formally yields two occupied cells and, subsequently, two non-zero entries in the binary occupancy vector $\boldsymbol{b}$. However, for relatively small deviations (depending on the spectral mask used) one of the cells contains a considerably smaller portion of the signal energy than the other. This means significantly different effective SNR within the cells. As a result, only the cell containing a larger part of the signal's energy is detected in the considered ISNR range, hence we observe a 50% recovery rate in the area around the optimal center frequency. At some point, the part of the signal's energy in the second cell becomes large enough to allow the effective SNR in both cells to approach the level required for correct detection. This, however, requires a larger input SNR, compared to the case of a single occupied cell, since the ISNR is independent of the signal sub-band position. The same mechanism applies for both C-DDR and Nyq-Ref; the difference between the two being that the latter tolerates lower ISNRs.

To evaluate average performance degradation due to the unknown sub-band location Figure 5.13 demonstrates the SRR averaged over 30 different sub-band locations where f_{c} is chosen uniformly at random from a set of LN values. We see that unknown sub-band location produces a significant overall performance deterioration, compared to the case when one it coincides with one of the spectral cells' locations. These results support our intuition (recall Example 5.18) that the signal sub-band splitting negatively affects the overall detection accuracy irrespective of the detection method used. Note, however, that if a presence of a small fraction of the active sub-band's energy (e.g., this happens when most of the energy belongs to one spectral cell and only an insignificant fraction of it to the neighbouring cell) gets undetected, it does not necessarily mean a significant negative effect on further processing.

5.4 Sparsity order estimation from sub-Nyquist samples

Having investigated the sub-Nyquist sampling framework and the approaches to multiband energy sensing from sub-Nyquist samples, we complete this chapter by taking one step back to the discussion on the sparsity order estimation (SOE) in Chapter 4 and showing how the SOE methods proposed there apply here.

5.4.1 SOE for multiband signals

As a a brief reminder, we begin with repeating the main expressions relating the wideband input $s(t)$ with the discrete narrowband outputs $\boldsymbol{y}[n]$. Thus,

$$\boldsymbol{y}(f) = \boldsymbol{A}\boldsymbol{z}(f), f \in \mathcal{F}_\mathrm{s}, \qquad \text{(rep. of (5.19))}$$

where the mth element of $\boldsymbol{y}(f)$ is the DTFT of the mth MWC output and $\boldsymbol{z}(f)$ is an unknown length-L vector with ℓth element

$$z_\ell(f) = S(f - \ell f_\mathrm{p}), f \in \mathcal{F}_\mathrm{s}. \qquad \text{(rep. of (5.20))}$$

In time domain, we have that

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{z}[n]. \qquad \text{(rep. of (5.22))}$$

5.4.1.1 Sparsity order of $\boldsymbol{z}[n]$

From (5.20), it is clear that only those entries of $\boldsymbol{z}(f \in \mathcal{F}_\mathrm{s})$ that correspond to the spectrum cells containing signal energy are non-zero. As discussed above, since $f_\mathrm{p} \geq B$ one narrowband bandpass signal $x_k(t)$ can potentially contribute only to up to two consecutive spectral cells. The total sparsity order of $\boldsymbol{z}(f)$, which we denote as K_z, is determined by the number of spectrum cells that are at least partially occupied and for a complex-valued multiband signal it is upper bounded[8] by $2K$ and lower bounded by K (i.e., $K \leq K_z \leq K$). Since $\boldsymbol{z}[n]$ is merely an IDTFT of $\boldsymbol{z}(f)$, only those of its elements are non-zero that correspond to the non-zero elements of $\boldsymbol{z}(f)$ and hence it is also K_z-sparse. Note that the sparsity order K_z depends not only on the number of signals K but also on the values of the central frequencies $f_{\mathrm{c},k}$ as well as the bandwidths B_k, as illustrated by Figures 5.4 and 5.12.

5.4.1.2 Observation covariance matrix analysis

Following (4.2), consider an observation covariance matrix $\boldsymbol{R}_\mathrm{y}$

$$\boldsymbol{R}_\mathrm{y} = \sum_{n=-\infty}^{\infty} \boldsymbol{y}[n]\boldsymbol{y}^\mathrm{H}[n] = \int_{f \in \mathcal{F}_\mathrm{s}} \boldsymbol{y}(f)\boldsymbol{y}^\mathrm{H}(f)\mathsf{d}f = \int_{f \in \mathcal{F}_\mathrm{s}} \boldsymbol{A}\boldsymbol{z}(f)\boldsymbol{z}^\mathrm{H}(f)\boldsymbol{A}^\mathrm{H}\mathsf{d}f = \boldsymbol{A}\boldsymbol{R}_\mathrm{z}\boldsymbol{A}^\mathrm{H}, \quad (5.50)$$

[8] Note that the sparsity order of $\boldsymbol{z}(f)$ with respect to each particular $f \in \mathcal{F}_\mathrm{s}$ is upper bounded by K since two parts of $X_k(f)$, when split into two spectral cells, do not overlap in $\mathcal{F}_\mathrm{s}$.

where $\boldsymbol{R}_{\mathrm{z}}$ is an $L \times L$ matrix with entries

$$[R_{\mathrm{z}}]_{k,j} = \int_{f \in \mathcal{F}_{\mathrm{s}}} z_k(f) z_j^*(f) \mathsf{d}f. \tag{5.51}$$

We established before that the diagonal elements of $\boldsymbol{R}_{\mathrm{z}}$ contain the energy of the corresponding spectrum cells and, therefore, $[\boldsymbol{R}_{\mathrm{z}}]_{j,j} \neq 0$ if and only if $j \in \mathcal{I}$, where $\mathcal{I}$ is defined by

$$\mathcal{I} = \bigcup_{i=1}^{K} \left\{ \left[\frac{\pm a_i}{f_s} \right] + L + 1, \left[\frac{\pm b_i}{f_s} \right] + L + 1 \right\}, \tag{5.52}$$

and $[\cdot]$ denotes rounding to the nearest integer. The non-diagonal elements $[\boldsymbol{R}_{\mathrm{z}}]_{k,j:k\neq j}$ represent cross-correlation terms between kth and jth spectral cells, where $[\boldsymbol{R}_{\mathrm{z}}]_{k,j} \neq 0$ if and only if $k, j \in \mathcal{I}$. Therefore, the rank of $\boldsymbol{R}_{\mathrm{z}}$ is equal to the cardinality of the index set $\mathcal{I}$ so that $\mathrm{rank}\,(\boldsymbol{R}_{\mathrm{z}}) = |\mathcal{I}| = K_{\mathrm{z}}$. Furthermore, it is shown in [202] that if the sensing matrix $\boldsymbol{A}$ has a Kruskal-rank[9] of at least K_{z}, then $\mathrm{rank}(\boldsymbol{R}_{\mathrm{y}}) = \mathrm{rank}(\boldsymbol{R}_{\mathrm{z}}) = K_{\mathrm{z}}$. This way the sparsity order K_{z} can be estimated as an effective rank of the observation covariance matrix $\boldsymbol{R}_{\mathrm{y}}$.

5.4.2 SOE with noisy observations

In practice, the wideband signal $s(t)$ is inevitably contaminated by a wideband signal noise $n_{\mathrm{w}}(t)$ at the receiver input and (possibly) by a narrowband measurement noise $\boldsymbol{n}_{\mathrm{m}}[n]$ at the sampler so that

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{z}[n] + \boldsymbol{n}_{\mathrm{z}}[n], \text{ where} \tag{5.53}$$

$$\boldsymbol{n}_{\mathrm{z}}[n] = \boldsymbol{A}\boldsymbol{n}_{\mathrm{s}}[n] + \boldsymbol{n}_{\mathrm{m}}[n]. \tag{5.54}$$

Here, $\boldsymbol{n}_{\mathrm{s}}[n]$ is an L-length vector with ℓth element being the IDTFT of the part of the wideband noise $n_{\mathrm{w}}(t)$ corresponding to the ℓth spectrum cell shifted to the origin. Taking into account (5.53) and (5.54), the covariance matrix of the noisy observations becomes

$$\boldsymbol{R}_{\mathrm{y}} = \boldsymbol{A}\boldsymbol{R}_{\mathrm{z}}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{R}_{\mathrm{n_z}} = \boldsymbol{A}\boldsymbol{R}_{\mathrm{z}}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{A}\boldsymbol{R}_{\mathrm{n_s}}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{R}_{\mathrm{n_m}}, \tag{5.55}$$

where $\boldsymbol{R}_{\mathrm{n_z}} = \boldsymbol{A}\boldsymbol{R}_{\mathrm{n_s}}\boldsymbol{A}^{\mathrm{H}} + \boldsymbol{R}_{\mathrm{n_m}}$ is the noise covariance matrix. The $\boldsymbol{R}_{\mathrm{n_m}}$ and $\boldsymbol{R}_{\mathrm{n_s}}$ in (5.55) denote the covariance matrices of $\boldsymbol{n}_{\mathrm{m}}[n]$ and $\boldsymbol{n}_{\mathrm{s}}[n]$, respectively.

At this point, the expression (5.55) is equivalent to the one derived for the discrete CS signal model considered in Chapter 4. Therefore, the sparsity order K_{z} can be estimated according to the same estimation routine. Namely, we first perform pre-whitening to the observations $\boldsymbol{y}[n]$ such that

$$\bar{\boldsymbol{y}}[n] = (\boldsymbol{R}_{\mathrm{n_z}})^{-1/2}\, \boldsymbol{y}[n], \tag{5.56}$$

[9] Since $\mathrm{krank}\,(\boldsymbol{A}) = \mathrm{spark}\,(\boldsymbol{A}) - 1$ [203], the requirement for the rank preservation is fulfilled whenever $\boldsymbol{A}$ conforms to the CS recovery condition as formulated by Theorem 1.

where $\boldsymbol{R}_{\mathrm{n_z}}$ should either be fully known at the receiver or has to be estimated beforehand, according to (4.9) for instance. Then, we estimate the effective rank of the sample observation covariance matrix $\hat{\boldsymbol{R}}_{\bar{\mathrm{y}}} = \frac{1}{T}\sum_{n=1}^{T} \bar{\boldsymbol{y}}[n]\bar{\boldsymbol{y}}^{\mathrm{H}}[n]$, $T \geq K_{\mathrm{z}}$, e.g., by applying MOS algorithms.

5.4.3 Numerical demonstration

To demonstrate SOE from sub-Nyquist sampled sparse multiband signals, we perform numerical simulations following a signal generation approach from previous section. Namely, we generate the test signals in the frequency domain by representing the continuous frequency band $\mathcal{F}$ by a grid of LN_{gr} equidistant points $f_n = nf_{\mathrm{d}}$. The multiband spectrum $S[f_n]$ is then computed as

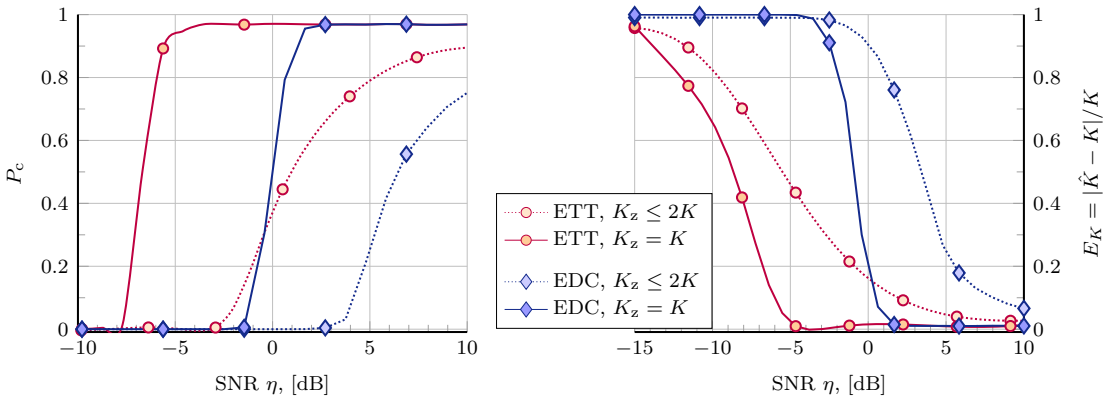
$$S[f_n] = \begin{cases} \sigma_{\mathrm{z},i} e^{\jmath\phi[f_n]} & f_n \in [a_i, b_i) \\ 0, & \text{otherwise} \end{cases}$$

where for each realization of $S[f_n]$

- $\phi[f_n]$ are drawn independently from a uniform distribution $\mathcal{U}(-\pi, \pi)$;
- $\sigma_{\mathrm{z},k}$ are drawn from a uniform distribution $\mathcal{U}(0, 1)$;
- $\{a_k, b_k\} = \{\langle f_{\mathrm{c},k} - B_k/2\rangle, \langle f_{\mathrm{c},k} + B_k/2\rangle\}$, where $f_{\mathrm{c},k} \in (B/2, W/2 - B/2)$ and $\langle\cdot\rangle$ denotes the operation of rounding to the nearest grid point.

The sensing matrix $\boldsymbol{A}$ is chosen randomly with entries $[\boldsymbol{A}]_{m,p}$ drawn independently from a complex normal distribution $\mathcal{CN}(0, 1)$ for $m \in [1, M]$ and $p \in [1, L_0 + 1]$, whereas for $p \in [L_0 + 2, 2L_0 + 1]$ $[\boldsymbol{A}]_{m,2L_0+1-p} = [\boldsymbol{A}]^*_{m,p}$. Both the wideband signal noise and the narrowband measurement noise are modelled as i.i.d. circularly symmetric Gaussian white noise vectors with elements with variance σ_{s}^2 and σ_{m}^2, respectively, where $\varsigma = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{m}}^2} = 1$. The total signal-to-noise ratio is defined as $\eta = \sum_i \sigma_{\mathrm{z},i}^2 / M\sigma_0^2$ with $\sigma_0^2 = \frac{1}{M}\mathrm{trace}\{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\}\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{m}}^2$. Additionally, in order to estimate the noise covariance matrix $\boldsymbol{R}_{\mathrm{n_z}}$ for the pre-whitening of the observations in (5.56), a training set of $T_{\mathrm{tr}} = 5 \cdot 10^3$ noise only observations is collected, whereas the number of time snapshots T used for both the sparsity order estimation and the simultaneous support recovery is set to 50 (resulting in $L_{\mathrm{tr}} = T_{\mathrm{tr}}/T = 100$). For the rank estimation from $\tilde{\boldsymbol{y}}[n]$ we use the ETT ($\alpha = 10^{-3}$) and the EDC.

Figure 5.14 presents the SOE results in terms of the probability of correct order estimation P_{c} (Figure 5.14a) and the normalized order estimation error $E_{K_{\mathrm{z}}}$ (Figure 5.14b) for the special case when each source signal $x_k(t)$ is confined within a single spectral cell ($K_{\mathrm{z}} = K$) as well as for the general case of arbitrary splitting of the source energy ($K_{\mathrm{z}} \leq 2K$). We see that the performance of the order estimation in the latter case is significantly inferior to the one for the former case. This is primarily due to the two following reasons. Firstly, the effective sparsity order in the case of no signal splitting is up to two times smaller than in the general case. Secondly and most importantly, when the signal sub-band is split between two neighbouring spectral cells its energy gets split as well, as discussed before. Subsequently,



(a) Probability of correct estimation

(b) Mean estimation error

Figure 5.14: SOE performance in the multiband scenario with respect to SNR: empirical probability of correct estimation P_c (a) and normalized sparsity order estimation error $E_K = \frac{\hat{K}-K}{K}$ (b). The results are presented for for the input signal with ($K_z \leq 2K$) and without ($K_z = K$) energy splitting.

although the total signal energy stays unchanged, it becomes distributed between larger number of eigenvalues which leads to a worse SNR per eigenvalue, which has an effect on the SOE performance similar to the one it had on the sensing performance.

5.5 Concluding remarks

In this chapter, we investigated the tasks of wideband multiband signal acquisition and spectrum sensing. As discussed in the beginning, there is a number of possible hardware solutions available for the reception of wideband multiband signals. They differ by a number of factors including simultaneous observed bandwidth, the number of RF chains, equivalent sampling rate, processing requirements, etc. Clearly, each solution has its advantages and disadvantages and is hence characterized by a distinct set of operational characteristics. This means that the choice of a particular architecture will likely depend on the specific application requirements at hand. Here, we have focused on a particular solution: the modulated wideband converter (MWC), which is a practical receiver architecture for obtaining sub-Nyquist samples of the wideband input. Its main advantages include simultaneous wideband reception at a reduced sampling rate, relatively straightforward hardware implementation compared to other A2I architectures (details on first MWC prototypes are available at [287]), reconfigurability as well as versatile application range. The main disadvantage of the MWC is the kind of restrictions it imposes on the types of signals it can operate with. As a receiver architecture that was developed to deal with signals belonging to sparse unions of subspaces, it is likely to be ill-fit for more generic scenarios and other signal classes. It also suffers from other negative effects that are generally common to compressive approaches, such as the noise folding discussed in Chapter 3. Nevertheless, the MWC presents one of the most promising solutions to date for sub-Nyquist sampling of sparse analog signals.

Having the MWC as the underlying receiver architecture, we studied approaches to coarse energy based multiband sensing. We showed that in the sub-Nyquist sampling framework one can directly recover signal's energy from sub-Nyquist samples at the frequency resolution of the mixing sequences, i.e., without the need to perform full signal or PSD recovery first. This allows detecting spectral occupancy within the consecutive sensing channels (implicitly) provided by the sampling hardware. Similar to PSD recovery, the proposed compressive coarse energy sensing does not require the spectral occupancy to be sparse, provided that the overall sampling rate is greater than half the Nyquist rate. However, further sampling rate reduction is only possible under the sparsity constraint. To evaluate the performance of such a direct compressive multi-channel energy detector (MED), we performed an extensive numerical study. For various signal scenarios we compared the empirical probabilities of correct detection, false alarm and missed detection, as well as the support recovery rate of the proposed approach to those of the related coarse wideband sensing methods based on multi-channel energy detection. Our results suggest that the direct coarse wideband sensing slightly outperforms its counterpart which operates on the recovered PSD, provided that the sensing parameters are equal. A further advantage of this approach comes from the fact that it can be used without setting an additional detection threshold. One should keep in mind however, that the spectral resolution of the proposed compressive energy detector is dictated by the parameters of the sampling scheme and hence it is fixed. This means that its use for fine WSS is only possible in conjunction with further spectrum/PSD reconstruction.

Finally, we explored the composition of the sensing matrix in MWC with respect to the spread of the effective signal energy at the output of the sampler, which proved to be largely in line with the results of our analysis of finite-dimensional CS systems in Chapter 3. We also demonstrated practical applicability of the rank-based sparsity order estimation from Chapter 4 in the multiband setting.

Chapter 6

Multiband spectrum sensing and DoA estimation

6.1 Motivation and related work

In the previous chapter, we discussed how using the modulated wideband converter (MWC) sub-Nyquist sampling architecture allows one to acquire an analog multiband signal at a low sampling rate without any loss of information, provided that its frequency support occupies only a part of the overall bandwidth. We also investigated approaches to multiband wideband spectrum sensing (WSS) from obtained low-rate sub-Nyquist samples with the aim of estimating the spectral occupancy of a wideband input across the observed frequency range. Along with the frequency support of a multiband signal, it is often desirable to estimate the direction of arrivals (DoAs) of the transmissions within the individual active sub-bands that constitute it. For example in cognitive radio (CR), having angular information about incoming transmissions provides a cognitive radio receiver with a *sense of direction* within its local neighborhood [238], [284]. This directional sense can then be exploited to help avoid interference to the primary system or to add a further resource domain. The ability to estimate DoAs in a multiband setting is also important to passive radar systems that use so-called "signals of opportunities" potentially originating from radio systems that operate in different frequency bands [291].

Generally, obtaining angular information requires the use of an antenna array. In view of compressive acquisition, in addition to an antenna one needs to devise a receiver structure for sub-Nyquist sampling and develop approaches for the reconstruction of angular and spectral sub-band locations from the received sub-Nyquist samples. It is worth noting that an application of sparse recovery and CS ideas to narrowband DoA estimation has attracted notable attention in the recent years and a variety of approaches for compressive wavefront sampling in the spatial domain has been described in the literature. A list of examples includes co-prime [292] and nested arrays [293], sparse random arrays [294]–[297],

and compressive arrays with linear combining networks [A3], [298], [299]. However, multiband frequency-angular sensing in a sub-Nyquist sampling framework is significantly less studied with only a few examples of possible receiver systems available. Thus, a method to estimate the angular-frequency power spectrum based on a sparsified uniform linear array (ULA) where each antenna is sampled at a sub-Nyquist rate using the multi-coset sampler (MCS) is discussed in [300]. Similarly, in [301] and [302] the authors propose antenna systems for sub-Nyquist acquisition of multiband signals with delay networks at the output of the antenna elements. Instead of placing an entire MCS at each antenna output, as in [300], they employ only a few additional delayed channels per antenna, depending on the considered antenna configuration. After the signal is sampled, the DoAs and the central frequencies of the individual transmissions are first estimated using MUSIC-based methods and then the frequency spectrum is reconstructed. An alternative approach is discussed in [303] where an L-shaped array of isotropic antenna elements is sampled at a sub-Nyquist rate by applying a single channel of the MWC at each antenna output. The central frequencies, together with the corresponding DoAs are jointly estimated by using either sparse recovery methods or classical methods for 2D-DoA estimation. The minimum number of antennas required for DoA estimation and blind spectrum reconstruction in [301]-[303] is shown to be proportional to the number of sources, yielding an overall minimal sampling rate consistent with [255]. Finally, the authors in [304] use the MWC channels to sample the outputs of a ULA at a sub-Nyquist rate, while estimating the DoAs and central frequencies of the signals within the active signal sub-bands separately. More specifically, they use the output of one antenna with a full MWC in order to estimate the central frequencies of the active sub-bands whereas the rest of the antenna outputs, sampled by employing a single MWC channel per antenna, are used to determine the DoAs. One of the main disadvantages of the methods described above is that they rely on specific antenna configurations, such as uniform linear or circular element placement, and assume that the individual elements have isotropic radiation patterns, which altogether implies a very particular structure of the antenna responses.

In this chapter, we generalize a sub-Nyquist sensing system based on the MWC to be applicable to (almost) arbitrary antenna arrays. We place several MWC channels at the output of each antenna element and develop recovery approaches that are largely independent of the structure of the antenna responses. More specifically, we apply a CS based approach assuming that the angles of arrival and central frequencies of the source signals lie on some predefined grid and derive conditions for perfect reconstruction of the angular-frequency spectrum in a noise-free environment. We demonstrate that using more MWC channels per antenna provides additional degrees of freedom for the system design, e.g., one can reduce the number of antennas provided that the source signals do not overlap in frequency. However, our numerical results indicate that in the presence of signal noise, this trade-off tends to be asymmetric favoring antennas over the MWC channels.

The material of this chapter is partially presented in [A11], while [A3], [A6] consider the task of compressive acquisition of RF signals for narrowband DoA estimation.

6.2 Multiband sub-Nyquist antenna system

6.2.1 Signal model

Input signal. Consider K independent far-field narrowband plane waves impinging on a wideband N-element antenna array. Each wave is associated with an unknown azimuth[1] angle of arrival θ_k and a baseband complex envelope $\bar{x}_k(t)$. The individual transmissions $x_k(t) = \bar{x}_k(t)e^{2\pi\jmath f_{\mathrm{c},k}t}$ are assumed to be uncorrelated, each of bandwidth $B_k \leq B$ and modulated to an unknown central frequency $f_k \in \mathcal{F} = [0, W)$. The signal impinging on the array can be then written as

$$x(t) = \sum_{k=1}^{K} x_k(t) = \sum_{k=1}^{K} \bar{x}_k(t)e^{2\pi\jmath f_{\mathrm{c},k}t}, \tag{6.1}$$

while the continuous angular-frequency spectrum of $x(t)$ is defined as

$$\mathcal{P}_x(\theta, f) = \sum_{k=1}^{K} X_k(f - f_{\mathrm{c},k})\delta(\theta - \theta_k), \tag{6.2}$$

where $X_k(f) = \int_{-\infty}^{+\infty} x_k(t)e^{-\jmath 2\pi f t}\mathsf{d}t$ is the Fourier transform of $x_k(t)$ and $\delta(\theta)$ is the Kronecker delta function. Then, we have that $\mathcal{P}_x(\theta, f) = 0, \forall f \notin \mathcal{F}$ and $\forall\theta \notin \{\theta_k\}_{k=1}^{K}$. An example of $\mathcal{P}_x(\theta, f)$ is illustrated in Figure 6.1.

Definition 9. *We define $\mathcal{X}$ as the set of all signals $x(t)$ that have structure defined by* (6.1)-(6.2) *such that the support of their angular-frequency spectrum is a union of K intervals in $\mathcal{F} = [0, W)$ of length $B_k \leq B$ that are located at K angles $\theta_k \in [-\pi, \pi]$.*

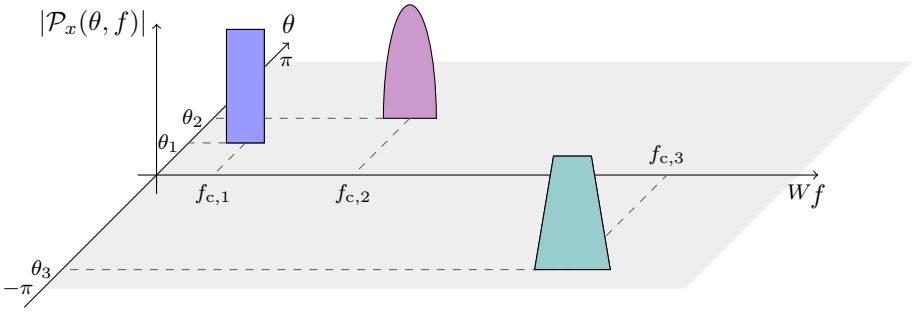


Figure 6.1: An example of an angular-frequency spectrum $\mathcal{P}_x(\theta, f)$ of a complex-valued multiband signal $x(t)$ comprised of $K = 3$ active sub-bands located at distinct central frequencies $f_{\mathrm{c},i} \in [0, W)$ with corresponding DoAs $\theta_k \in [-\pi, \pi]$.

[1]For the sake of brevity, we assume here that the sources are located in the azimuthal plane of the array and the impinging waves are co-polarized with it, while an extension to a more general case is outlined in Section 6.4.

Received signal. Under the narrowband assumption on the individual signals $x_k(t)$ the Fourier transform of the signal $s_n(t)$ received at nth antenna port can be expressed as

$$S_n(f) = \sum_{k=1}^{K} a_n(\theta_k, f_{\mathrm{c},k}) \mathcal{P}_x(\theta_k, f), \tag{6.3}$$

where $a_n(\theta, f)$ is the response of the n-th antenna element as a function of the azimuth[2] angle θ and frequency f. For example, for a uniform circular array (UCA) with isotropic elements we know that

$$a_n(\theta_k, f_{\mathrm{c},k}) = e^{j2\pi R/\lambda_k cos(\theta_k - \vartheta_n)}, \tag{6.4}$$

where $\lambda_k = c/f_k$ with $c = 3 \cdot 10^8$ m/c, $\vartheta_n = 2\pi(n-1)/N$ and R being the array radius. A collection of the responses from all N antenna elements for some pair $(\theta_k, f_{\mathrm{c},k})$ constitutes an array steering vector $\boldsymbol{a}(\theta_k, f_k) = [a_1(\theta_k, f_k), a_2(\theta_k, f_k), \ldots, a_N(\theta_k, f_k)]^{\mathrm{T}}$, whereas by $\boldsymbol{a}(\theta, f)$ we denote the array manifold as a function of a (continuous) angle θ and frequency f.

Definition 10. *An array manifold $\boldsymbol{a}(\theta, f)$ of some N-element antenna array is called ambiguity-free if any K distinct array steering vectors $\boldsymbol{a}(\theta_k, f_k) \in \boldsymbol{a}(\theta, f)$, $k = 1, 2, \ldots, K$ are linearly independent for all $K \leq N$ and $f_k \in \mathcal{F}$ [305].*

It is worth noting that for a manifold of an array with isotropic elements to be ambiguity-free it is required that $N > 2$. This is due to the fact that the phase offset between any two isotropic antenna elements is $2\pi d/cf \cos(\theta)$, where d is the distance between the elements. Thus, for any $N = 2$ elements there always exists a distinct pair $\{(\theta_j, f_i), (\theta_p, f_q)\}$ such that $\boldsymbol{a}(\theta_j, f_i) = \boldsymbol{a}(\theta_p, f_q)$. The ambiguity-free property has been shown analytically to be fulfilled for some special cases, such as narrowband uniform linear or circular arrays [306]. However, in general it is difficult to verify that a given array is ambiguity-free. At the same time, numerical evidence suggests that this is not a major concern in the wideband case since encountering linear dependencies in practically built arrays is highly unlikely.

Definition 11. *The set $\mathcal{A}$ contains all N-element antenna arrays that satisfy the following assumptions: (A1) the array manifold is ambiguity-free in a sense of Definition 10; (A2) the spatial-spectral response of every port $\boldsymbol{a}(\theta, f)$ is non-zero almost everywhere.*

In the following we describe a receiver system for sub-Nyquist sampling of $s_n(t)$ that allows for perfect reconstruction of $\mathcal{P}_x(\theta, f)$ for any $x(t)$ in $\mathcal{X}$ and any antenna with array manifold $\boldsymbol{a}(\theta, f)$ in $\mathcal{A}$, while reducing the sampling rate in each channel below what is required by the Nyquist sampling theorem.

[2]Note that, as mentioned above, we consider here a special case when the array response depends on the azimuth angle θ only. In general, in addition to frequency and azimuth angle of arrival, a_n is also a function of the elevation angle of arrival and the polarization state of the incident plain wave.

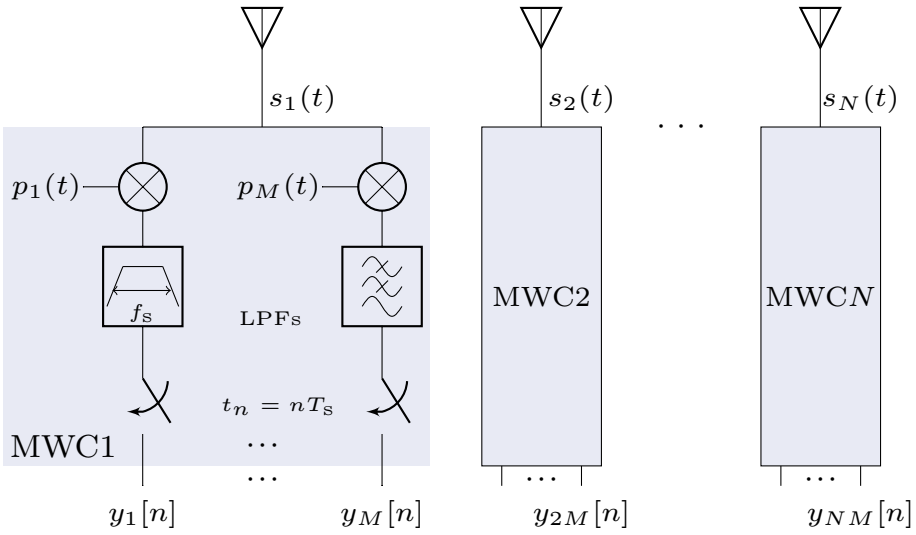


Figure 6.2: Block diagram of a multiband antenna system with sub-Nyquist sampling that consists of N antenna elements, each followed by an M-channel MWC.

6.2.2 Sub-Nyquist receiver system

6.2.2.1 System description

Consider a receiver architecture consisting of N antenna elements followed by N MWCs, as shown in Figure 6.2. From Section 5.3.1 we know that in MWC the input signal is fed into M sampling channels where it is multiplied by a T_p-periodic pseudo-random sequence $p_i(t)$, low-pass filtered and sampled at a sampling rate[3] of f_s with $f_s = f_p = 1/T_p \geq B$ [255]. In a multiple antenna system, the total number of sampling channels is $M_t = NM$ while the total sampling rate is equal to $M_t f_s$. Generally, the sets of mixing sequences $\{p_i(t)\}_{i=(n-1)M+1}^{nM}$, $n = 1, 2, \ldots, N$ can be different for different antenna elements. Hence, the received signal in the mth sampling channel of the nth antenna element is multiplied by the periodic sequence $p_d(t)$, where $d = (n-1)M + m$ and, due to periodicity,

$$p_d(t) = \sum_{\ell=-\infty}^{\infty} c_{d,\ell} e^{\jmath 2\pi \ell f_p t}, \text{ with } c_{d,\ell} = \frac{1}{T_p} \int_0^{T_p} p_d(t) e^{-\jmath 2\pi \ell f_p t} \mathsf{d}. \tag{6.5}$$

6.2.2.2 Frequency domain analysis

Consider the Fourier transform of the analog multiplication $\tilde{y}_d(t) = s_n(t) p_d(t)$, i.e.,

$$\tilde{y}_d(f) = \sum_{\ell=-\infty}^{\infty} c_{d,\ell} S_n(f - \ell f_p). \tag{6.6}$$

[3] Although, the sampling rate in MWC can be greater than f_p (which results in the expander structure (5.28)), for the sake of simplicity in the following we consider the basic configuration of $f_s = f_p$.

Substituting (6.3) into (6.6) and taking into account that the input signal $x(t)$ is bandlimited, we get the following expression for the DTFT of the dth receiver output after sampling

$$y_d(e^{2\pi f T_\mathrm{s}}) = \sum_{\ell=-L_0}^{L_0} c_{d,\ell} \sum_{k=1}^{K} a_n(\theta_k, f_{\mathrm{c},k}) \mathcal{P}_x(\theta_k, f - \ell f_\mathrm{p}),\ f \in \mathcal{F}_s = [-f_\mathrm{s}/2, f_\mathrm{s}/2], \tag{6.7}$$

where $L_0 = \left\lceil \frac{f_\mathrm{Nyq}+f_\mathrm{s}}{2f_\mathrm{p}} \right\rceil - 1$ and $f_\mathrm{Nyq} = W$. Denote now by $\boldsymbol{y}_n(f)$ the vector collecting M DTFTs of the outputs of nth antenna, i.e., jth element of $\boldsymbol{y}_n(f)$ is given by $y_{n,j}(f) = y_{(n-1)M+j}(e^{2\pi f T_\mathrm{s}})$. Then, we have that

$$\boldsymbol{y}_n(f) = \boldsymbol{\Phi}_n \sum_{k=1}^{K} a_n(\theta_k, f_{\mathrm{c},k}) \boldsymbol{z}_k(f), f \in \mathcal{F}_s, \tag{6.8}$$

where $\boldsymbol{\Phi}_n$ is an $M \times L = 2L_0+1$ matrix with (m,j)th element $[\boldsymbol{\Phi}_n]_{m,j} = c_{d,j-L_0-1} = c^*_{d,L_0+1-j}$. The unknown vector $\boldsymbol{z}_k(f)$ contains ℓf_p-shifted low-passed filtered copies of $\mathcal{P}_x(\theta_k, f)$ with

$$z_{k,j}(f) = \mathcal{P}_x(\theta_k, f - \ell f_\mathrm{p}),\ f \in \mathcal{F}_\mathrm{s}, \tag{6.9}$$

By concatenating $\boldsymbol{y}_n(f)$ into one length-NM vector $\boldsymbol{y}(f)$, we obtain

$$\boldsymbol{y}(f) = \boldsymbol{\Psi} \boldsymbol{z}(f),\ f \in \mathcal{F}_\mathrm{s}. \tag{6.10}$$

Here, the unknown vector $\boldsymbol{z}(f)$ is comprised of $\boldsymbol{z}_k(f)$ stacked one under another and $\boldsymbol{\Psi}$ is an $NM \times LK$ matrix with (d,g)th entry given by

$$[\boldsymbol{\Psi}]_{d,g} = c_{d,\ell} a_n(\theta_k, f_{\mathrm{c},k}), \tag{6.11}$$

where $g = (k-1)L + L_0 + \ell + 1$. Note that, similarly to $\boldsymbol{z}(f)$ in (5.19) the elements of $\boldsymbol{z}_k$ contain spectrum slices of $\mathcal{P}_x(\theta_k, f)$ of width f_s and central frequencies ℓf_p where $f_\mathrm{p} \geq B$. Thus each $\mathcal{P}_x(\theta_k, f)$ contributes only one non-zero element to $\boldsymbol{z}_k(f)$ for any fixed $f \in \mathcal{F}_\mathrm{s}$ and the vector $\boldsymbol{z}(f)$ contains at most K non-zeros.

Finally, we can also write (6.10) in the time domain as

$$\boldsymbol{y}[n] = \boldsymbol{\Psi} \boldsymbol{z}[n], \tag{6.12}$$

where $\boldsymbol{y}[n] = [y_1[n], \ldots, y_{NM}[n]]^\mathrm{T}$ is the receiver output, as indicated in Figure 6.2, while $\boldsymbol{z}[n]$ is the IDTFT of $\boldsymbol{z}(f)$. Note that (6.10), (6.12) are conceptually similar to (5.19), (5.22); they differ by the composition of the dictionary $\boldsymbol{\Psi}$ and the exact structure of $\boldsymbol{z}(f)$, $\boldsymbol{z}[n]$.

Expressions (6.9)–(6.12) bind the angular-frequency spectrum of the wideband input $x(t)$ to the DTFT of the low rate output $\boldsymbol{y}[n]$. In the following, we present conditions for recovery of $\boldsymbol{z}(f)$ from (6.10) and provide concrete methods to estimate $\{\theta_k\}_{k=1}^K$, $\{f_{\mathrm{c},k}\}_{k=1}^K$.

6.3 Reconstruction approaches

In this section, we develop approaches for estimation of the central frequencies and the associated DoAs from (6.10). We assume a grid-based approach where the antenna manifold $\boldsymbol{a}(\theta, f)$ is represented by an $N_\theta \times N_f$ grid with N_θ points in angle and N_f in frequency.

6.3.1 Minimal sampling rate

Denote by $\theta_\mathrm{d} = 2\pi/N_\theta$ and $f_\mathrm{d} = f_\mathrm{Nyq}/N_f$ grid steps in azimuth and frequency, respectively, and suppose that for all $k = 1, 2, \dots K$ the pairs $(\theta_k, f_{\mathrm{c},j})$ lie on the grid $\{\alpha\theta_\mathrm{d}\}_{\alpha=1}^{N_\theta} \times \{\beta f_\mathrm{d}\}_{\beta=0^{N_f}}$, where $\times$ denotes the Cartesian set product. Then, we can write (6.10) as

$$\boldsymbol{Y}(f) = \boldsymbol{B}\boldsymbol{w}(f),\ f \in \mathcal{F}_\mathrm{s}, \tag{6.13}$$

where $\boldsymbol{B}$ is an $NM \times N_\theta N_f L$ matrix with its (d,p)th element given by

$$[\boldsymbol{B}]_{d,p} = c_{d,\ell} a_n(\alpha\theta_\mathrm{d}, \beta f_\mathrm{d}). \tag{6.14}$$

Here, the index d is the same as in (6.6), namely $d = (n-1)M + m$, while p is the grid index defined by $\alpha \in [1, N_\theta]$, $\beta \in [0, N_f]$) and $\ell \in [-L_0, L_0]$ as

$$p = (\alpha - 1)N_f L + \beta L + L_0 + \ell + 1. \tag{6.15}$$

Thus, the matrix $\boldsymbol{B}$ in (6.13) is an expanded version of $\boldsymbol{\Psi}$, i.e., $\boldsymbol{\Psi} = \boldsymbol{B}_{\bigcup_k \mathcal{I}_k}(f)$ where $\mathcal{I}_k = \{p : \alpha = \theta_k/\theta_\mathrm{d}, \beta = f_k/f_\mathrm{d}, \ell = -L_0, \dots, L_0\}$ is the index set of active atoms in the expanded dictionary $\boldsymbol{B}$. Accordingly, the non-zero elements of the sparse vector $\boldsymbol{w}(f)$ are the non-zeros of $\boldsymbol{z}(f)$ where $\boldsymbol{z}_k(f) = \boldsymbol{w}_{\mathcal{I}_k}(f)$.

Denote by $1 \leq \mu \leq K$ the maximum number of distinct indices ℓ corresponding to the positions of the non-zeros[4] in $\boldsymbol{w}(f)$ and consider the following theorem.

Theorem 7. *Let $\boldsymbol{a}(\theta, f)$ be an arbitrary array manifold in $\mathcal{A}$ and $x(t)$ be an arbitrary signal in $\mathcal{X}$ sampled according to Figure 6.2 with $f_\mathrm{s} = f_\mathrm{p} \approx B$. In a noiseless environment, the minimal total number of channels $M_\mathrm{t} = NM$ required for perfect recovery of $\boldsymbol{w}(f)$ in (6.13) for any $f \in \mathcal{F}_\mathrm{s}$ is $M_\mathrm{t} \geq 2K$ with $N > K - \min(M, \mu) + 1$, irrespective of $\boldsymbol{\Phi}_n$ as long as it has full Kruskal-rank for any $n = 1, 2, \dots N$.*

Proof. The proof uses the fact that (6.13) has a unique solution if $\mathrm{spark}(\boldsymbol{B}) > 2K$ [10]. The conditions of the theorem then follow from showing that by choosing appropriate $\boldsymbol{\Phi}_n$ it is always possible to ensure that any $r \leq 2K$-column submatrix of $\boldsymbol{B}$ is full rank, as explained in details in Appendix B.4.1. ■

[4]Note that when $\mu < K$ it means that several transmissions (partially) occupy the same spectral cell while having different DoAs and/or central frequencies.

Theorem 7 states that the minimum required number of channels in the proposed system is $2K$ resulting in the minimum sampling rate of $2KB$. Additionally, it establishes the conditions under which $M > 1$ MWC channels can be used to reduce the number of antennas. Thus, if none of the frequency supports of the source signals overlap, i.e., $\mu = K$, the values of N and M can be chosen arbitrary such that their product is greater than $2K$ while $N > 2$. However, the more sources occupy a spectrum cell with the same index the fewer degrees of freedom there are. In the extreme case where all sources share the same spectrum cell ($\mu = 1$), we can recover only $N/2$ sources regardless of M. The intuition behind this observation is that while the antenna array is frequency- and angle-dependent, the MWC is only frequency-dependent, which means that it is not capable of helping resolving signal sources that overlap in frequency. Furthermore, Theorem 7 establishes that the matrices $\boldsymbol{\Phi}_n$ can be identical such that $\boldsymbol{\Phi}_n = \boldsymbol{\Phi}\ \forall n = 1, 2, \dots N$ as long as $\boldsymbol{\Phi}$ is full Kruskal-rank. This means that each antenna can have exactly the same MWC sampling block. It is also worth noting that Theorem 7 does not impose any specific requirements on the structure of the antenna responses but the conditions from Definition 11, which are generally desired features of any antenna system [307]. This being said, the proposed framework even allows incorporating measured antennas, by applying a measurement-based model according to the effective aperture distribution function (EADF) [314] for instance.

6.3.2 Parameter estimation

From (6.14) and (6.15), we have that the values of the elements in $\boldsymbol{\theta}$ and $\boldsymbol{f}$ are encoded in the indices of the active atoms of $\boldsymbol{B}$ that, in turn, correspond to the indices of the non-zero elements of $\boldsymbol{w}(f)$. Therefore, estimating the support of $\boldsymbol{w}(f)$ is tantamount to estimating the central frequencies and the associated DoAs of the K active signal sources. In the following, we present several possible approaches to do so. In the first two, we recover the frequency-angular support jointly, while in the last one we apply an approach similar to that from [304] where we first estimate the frequency support of $\boldsymbol{w}(f)$, after which we determine θ_k and $f_{\mathrm{c},k}$.

6.3.2.1 Joint support recovery

Similarly to (5.19), the linear system (6.13) contains an infinite number of equations. To solve it, we can either apply the CTF block from Section 2.4 or the energy based approach from Section 5.3.2, as described below.

CTF block. The first method we consider is the CTF block based support recovery described in Section 2.4, which we applied to spectral recovery in Section 5.3.1.4. Consider an $NM \times NM$ covariance matrix $\boldsymbol{Q}$ defined as

$$\boldsymbol{Q} = \sum_{n=-\infty}^{\infty} \boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n] = \int_{f \in \mathcal{F}_{\mathrm{s}}} \boldsymbol{y}(f)\boldsymbol{y}^{\mathrm{H}}(f)\mathsf{d}f. \tag{6.16}$$

Substituting (6.13) in (6.16), we have that

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{R}_{\mathrm{w}}\boldsymbol{B}^{\mathrm{H}}, \tag{6.17}$$

where $\boldsymbol{R}_{\mathrm{w}} = \int\limits_{f\in\mathcal{F}_s} \boldsymbol{w}(f)\boldsymbol{w}^{\mathrm{H}}(f)\mathsf{d}f$ is an $N_\theta N_f L \times N_\theta N_f L$ input signal covariance matrix. Given $\boldsymbol{Q}$, we can construct a frame $\boldsymbol{V}$ such that $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}$, e.g., by performing eigenvalue decomposition of $\boldsymbol{Q}$ and choosing $\boldsymbol{V}$ as its eigenvectors [36]. Once $\boldsymbol{V}$ is found, we solve the following system for the sparsest $\boldsymbol{U}$

$$\boldsymbol{V} = \boldsymbol{B}\boldsymbol{U}. \tag{6.18}$$

As before, the support of the unique solution to (6.18) is equal to the joint support of the unique solution to (6.13). Since (6.18) is a typical MMV problem, it can be solved[5] by any of the available MMV sparse recovery algorithms, such as the SOMP from [134] for instance.

Energy recovery. To find the support of $\boldsymbol{w}(f)$, we could also use the fact the individual signals $x_k(t)$ are independent from each other and apply the energy detection strategy presented in Section 5.3.2. More specifically, the covariance matrix $\boldsymbol{R}$ in (6.17) is a diagonal matrix with at most $2K$ non-zero elements on its main diagonal. Therefore, by vectorizing $\boldsymbol{Q}$ we obtain

$$\boldsymbol{q} = \mathrm{vec}(\boldsymbol{Q}) = (\boldsymbol{B}^* \otimes \boldsymbol{B})\,\boldsymbol{r}_{\mathrm{w}} = (\boldsymbol{B}^* \odot \boldsymbol{B})\,\boldsymbol{\epsilon}, \tag{6.19}$$

where $\boldsymbol{r}_{\mathrm{w}} = \mathrm{vec}(\boldsymbol{R}_{\mathrm{w}})$ while $\boldsymbol{\epsilon} = \mathrm{diag}\,(\boldsymbol{R}_{\mathrm{w}})$. Obviously, the support of $\boldsymbol{\epsilon}$ coincides with that of $\boldsymbol{w}(f \in \mathcal{F}_{\mathrm{s}})$ and, hence, recovering $\boldsymbol{\epsilon}$ from $\boldsymbol{q}$ we obtain the indices of the active atoms in $\boldsymbol{B}$.

Once the positions $p_i \in \mathcal{S}(\boldsymbol{w}(f \in \mathcal{F}_{\mathrm{s}}))$ of the non-zero elements of $\boldsymbol{w}(f)$ are found, the active angles and central frequencies are given by

$$\hat{\theta}_i = \left\lceil \frac{p_i}{N_f L} \right\rceil \theta_{\mathrm{d}}, \tag{6.20}$$

$$\hat{f}_{\mathrm{c},i} = \left\lceil \frac{p_i - (\frac{\hat{\theta}_i}{\theta_{\mathrm{d}}} - 1)N_f L}{L} \right\rceil f_{\mathrm{d}}, \tag{6.21}$$

with $i = 1, 2, \ldots, |\mathcal{S}(\boldsymbol{w}(f \in \mathcal{F}_{\mathrm{s}}))|$. From $\mathcal{S}(\boldsymbol{w}(f \in \mathcal{F}_{\mathrm{s}}))$ we can also obtain the indices of the active spectral cells as

$$\hat{\ell}_i = \left\lceil p_i - \left(\frac{\hat{\theta}_i}{\theta_{\mathrm{d}}} - 1\right) N_f L - \left(\frac{\hat{f_{\mathrm{c},i}}}{f_{\mathrm{d}}} - 1\right) L \right\rceil, p_i \in \mathcal{S}. \tag{6.22}$$

[5]Note that solving (6.18) generally requires twice the minimal sampling rate, i.e., $4KB$. This is due to the fact that although $\boldsymbol{W}(f)$ is K-sparse with respect to any fixed $f \in \mathcal{F}_{\mathrm{s}}$, the maximum joint sparsity of $\boldsymbol{W}(f \in \mathcal{F}_{\mathrm{s}})$ is $2K$ as each transmission can be split between up to two adjacent spectral cells.

6.3.2.2 Two-step support recovery

The main disadvantage of solving (6.18) or (6.19) is the size of the recovery problem. Note that $\boldsymbol{B}$ is constructed by detailing a system response over the azimuth angle θ, central frequency f_{c} and the spectral cell index ℓ. Depending on the grid resolution in frequency and angle, this can yield a significant computational complexity. To reduce the search space, we could recover the indices of the active spectral cells from the outputs of one antenna first, as in regular MWC, and then use a reduced-size dictionary to recover $\boldsymbol{\theta}$, $\boldsymbol{f}$. To be more specific, consider the outputs of first antenna $\boldsymbol{y}_1[n] = [y_1[n], y_2[n], \ldots, y[n]_M]^{\mathrm{T}}$. Viewed in frequency domain, it is represented by (6.8), which we can re-write as

$$\boldsymbol{y}_1(f) = \boldsymbol{\Phi}_1 \sum_{k=1}^{K} a_1(\theta_k, f_{\mathrm{c},k}) \boldsymbol{z}_k(f) = \boldsymbol{\Phi}_1 \boldsymbol{z}(f), f \in \mathcal{F}_s, \tag{6.23}$$

where $\boldsymbol{z}(f)$ is an length-L vector that is at most $2K$-sparse. Note that (6.23) is formally identical to (5.19) but for one crucial difference, namely that we do not require here that the individual signals occupy distinct spectral cells. This means that the cardinality of the support set $\mathcal{S}(\boldsymbol{z}(f \in \mathcal{F}_{\mathrm{s}}))$ is lower bounded by 1 rather than by K, as in the original MWC application. Nevertheless, what we are interested in is the indices of the active spectral cells within $\boldsymbol{z}(f)$. In light of this, we denote the support of $\boldsymbol{z}(f)$ as $\mathcal{I}_\ell = \mathcal{S}(\boldsymbol{z}(f \in \mathcal{F}_{\mathrm{s}}))$.

Once the index set $\mathcal{I}_\ell$ is known, e.g., by applying the recovery strategy from previous chapter, we can reduce (6.13) to

$$\boldsymbol{y}(f) = \tilde{\boldsymbol{B}} \tilde{\boldsymbol{w}}(f), \tag{6.24}$$

where $\tilde{\boldsymbol{w}}(f)$ is a vector of length $N_\theta N_f |\mathcal{I}_\ell| \ll N_\theta N_f L$ that contains the elements of $\boldsymbol{w}(f)$ corresponding to the spectral cell indices in $\mathcal{I}_\ell$, i.e., $\tilde{\boldsymbol{w}}(f) = \boldsymbol{w}_{\tilde{\mathcal{I}}_\ell}$ where $\tilde{\mathcal{I}}_\ell = \{(\alpha-1)N_f L + \beta L + L_0 + \ell_i + 1\}_{\ell_i \in \mathcal{I}_\ell}$. Accordingly, the matrix $\tilde{\boldsymbol{B}}$ in (6.24) is a submatrix of $\boldsymbol{B}$ of size $NM \times N_\theta N_f |\mathcal{I}_\ell|$ such that $\tilde{\boldsymbol{B}} = \boldsymbol{B}_{\tilde{\mathcal{I}}_\ell}$. The rest follows as before: solving (6.24) we obtain the support of $\tilde{\boldsymbol{w}}(f)$, which in turn allows us to estimate the elements of $\boldsymbol{\theta}$, $\boldsymbol{f}$ as

$$\begin{aligned} \hat{\theta}_i &= \left\lceil \frac{p_i}{N_f |\mathcal{I}_\ell|} \right\rceil \theta_{\mathrm{d}}, \\ \hat{f}_{\mathrm{c},i} &= \left\lceil \frac{p_i - (\frac{\hat{\theta}_i}{\theta_{\mathrm{d}}} - 1) N_f |\mathcal{I}_\ell|}{|\mathcal{I}_\ell|} \right\rceil f_{\mathrm{d}}, \end{aligned} \tag{6.25}$$

where $p_i \in \mathcal{S}(\tilde{\boldsymbol{w}}(f \in \mathcal{F}_{\mathrm{s}}))$ and $i = 1, 2, \ldots, |\mathcal{S}(\tilde{\boldsymbol{w}}(f \in \mathcal{F}_{\mathrm{s}}))|$.

Note that the minimal sampling rate here remains dictated by Theorem 7. What solving (6.23)–(6.24) brings is the reduction of the dictionary size and, subsequently, the reduction of the search space for parameter estimation. On the other hand, if the frequency support in (6.23) is estimated incorrectly, then the recovery in (6.24) fails almost surely.

6.3.3 2D spectrum reconstruction

Knowing the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{f}$, we might also be interested in reconstructing $\mathcal{P}_x(\theta, f)$. Before we proceed let us, for simplicity, denote $\mathcal{S}(\boldsymbol{w}(f \in \mathcal{F}_\mathrm{s}))$ by $\mathcal{S}$ and $\mathcal{S}(\tilde{\boldsymbol{w}}(f \in \mathcal{F}_\mathrm{s}))$ by $\tilde{\mathcal{S}}$. To reconstruct $\mathcal{P}_x(\theta, f)$, we write (6.13) as

$$\boldsymbol{y}(f) = \boldsymbol{B}_\mathcal{S} \boldsymbol{w}_\mathcal{S}(f), \tag{6.26}$$

where $\boldsymbol{B}_\mathcal{S}$ and $\boldsymbol{w}_\mathcal{S}$ contain the columns of $\boldsymbol{B}$ and the entries of $\boldsymbol{w}$ indexed by $\mathcal{S}$, respectively. From the time-domain samples $\boldsymbol{y}[n]$, we can now estimate the IDTFT of $\boldsymbol{w}_\mathcal{S}(f)$ as

$$\boldsymbol{w}_\mathcal{S}[n] = \boldsymbol{B}_\mathcal{S}^\dagger \boldsymbol{y}[n]. \tag{6.27}$$

In the case of the two-step approach, we re-write (6.24) as $\boldsymbol{Y}(f) = \boldsymbol{B}_{\tilde{\mathcal{S}}} \boldsymbol{W}_{\tilde{\mathcal{S}}}(f)$ and obtain $\boldsymbol{w}_{\tilde{\mathcal{S}}}[n]$ instead. The following theorem provides sufficient conditions for the recovery of $\mathcal{P}_x(\theta, f)$ from $\boldsymbol{y}[n]$.

Theorem 8. *Let $\hat{\boldsymbol{w}}(f)$ be the unique solution to (6.13). If $f_\mathrm{s} = f_\mathrm{p} \geq B$ and $\boldsymbol{\Phi}_n$ has full Kruskal rank for any $n = 1, 2, \ldots, N$, then $\mathcal{P}_x(\theta, f)$ can be uniquely recovered from $\boldsymbol{y}[n]$.*

Proof. The proof follows along the same lines as that of Theorem 2 from [255] taking into account the fact that for $\boldsymbol{\Phi}_n$ with full Kruskal rank $\boldsymbol{B}$ is invertible, which in turn follows from the proof of Theorem 7. ■

Under the conditions of Theorem 8, the angular-frequency spectrum $\mathcal{P}_x(\theta, f)$ can be uniquely reconstructed as

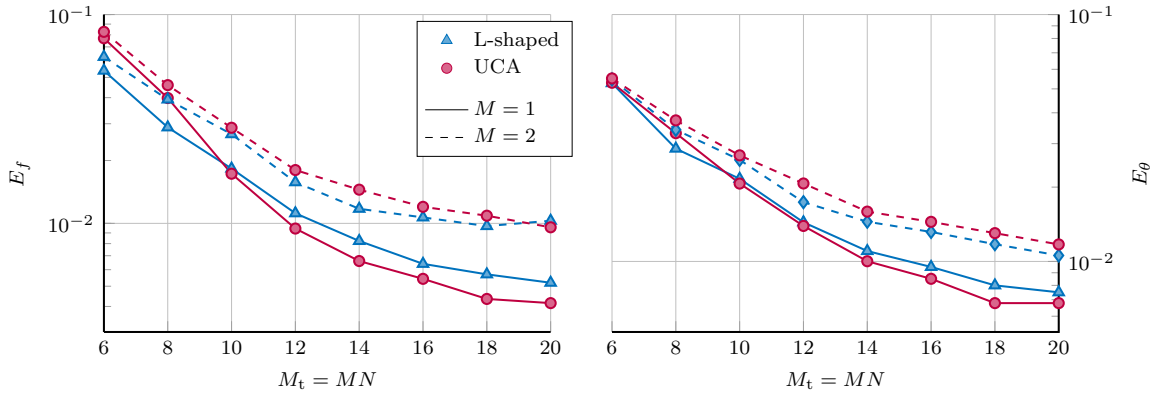
$$\hat{\mathcal{P}}_x(\theta, f) = \sum_i W_{p_i}(e^{\jmath 2\pi (f + \hat{\ell}_i f_p) T_s}) \delta(\theta - \hat{\theta}_i), \tag{6.28}$$

where $w_{p_i}(e^{\jmath 2\pi f T s})$ is the DTFT of $w_{p_i}[n]$ and $p_i \in \mathcal{S}$.

6.3.4 Performance evaluation

6.3.4.1 Simulation setup

In order to verify the proposed antenna system, we simulate a scenario with $K = 2$ uncorrelated signals $x_i(t)$ of equal magnitude and bandwidth $B = 10$ MHz. Each signal is modulated to an unknown central frequency $f_k \in [0, f_\mathrm{Nyq}/B]$ with $f_\mathrm{Nyq} = 2$ GHz and is associated with an unknown DoA $\theta_k \in [-\pi/2, \pi/2]$. The central frequencies $f_{\mathrm{c},k}$ and the DoAs θ_k lie on uniform grids of $N_\theta = 45$ and $N_F = 153$ points, respectively. The signal $x(t) = \sum_{k=1}^K \bar{x}_k(t) e^{2\pi \jmath f_k t}$ is received by an N-element antenna array according to Figure 6.2 with $f_\mathrm{p} = f_\mathrm{s} = 1.3B$. For mixing we use $p_m(t) = \sum_\ell \delta(t - \ell T_\mathrm{p} - \tau_m T_\mathrm{Nyq})$ with $T_\mathrm{Nyq} = 1/f_\mathrm{Nyq}$ and $\tau_m \in \mathbb{Z}$ being generated uniformly at random in the interval $[1, \frac{T_\mathrm{p}}{T_\mathrm{Nyq}} - 1]$. The antenna is



(a) carrier frequency esimtation error **(b)** DoA esimtation error

Figure 6.3: Average carrier frequency (a) and DoA (b) estimation error vs. the number of output channels $M_\mathrm{t} = MN$ for an SNR of 20 dB.

modeled as: 1) uniform circular array (UCA) with isotropic elements positioned equidistantly on a circle with a radius of $R = \frac{N}{10}\frac{2c}{f_\mathrm{Nyq}}$; 2) L-shaped array formed by two perpendicularly oriented uniform linear arrays (ULA) with $N/2$ isotropic elements each and the distance between the elements $d = \frac{c}{f_\mathrm{Nyq}}$. In order to solve the CTF block, we apply the SOMP algorithm from [134], where $\boldsymbol{Q}$ is estimated from 600 time samples of $\boldsymbol{y}[n]$, while all the following results are obtained by averaging the respective values over 200 realizations. Note that below we provide the results for the joint estimation approach only.

6.3.4.2 Parameter estimation accuracy

We begin our numerical investigation by examining the support recovery performance with respect to the joint estimation of the DoAs and the central frequencies. Similarly to [308], we use the following performance metrics: the normalized central frequency estimation error E_f and the normalized DoA estimation error E_θ defined as

$$E_f = \frac{\frac{1}{K}\sum_{k=1}^{K}\left|f_{\mathrm{c},k} - \hat{f}_{\mathrm{c},k}\right|}{f_\mathrm{Nyq}}, \text{ and } E_\theta = \frac{\frac{1}{K}\sum_{k=1}^{K}\left|\theta_k - \hat{\theta}_k\right|}{\pi},$$

respectively. Figure 6.3 presents the average DoA and central frequency estimation errors depending on the total number of output channels $M_\mathrm{t} = MN$ for two values of M, both considered antenna configurations and an SNR of 20 dB. Note that when $M = 2$ the number of antennas N is twice smaller than the total number of receiver outputs, e.g., for $M_\mathrm{t} = 10$ and $M = 2$ we have that the total of $N = 5$ antennas are used, each with 2 MWC channels.

Presented results suggest that both antenna configurations provide largely comparable performance with respect to the parameter estimation accuracy. In the noisy environment however, the trade-off between the number of antennas N and the number of sampling channels per antenna M tends to be asymmetric, which manifests itself in a slight but

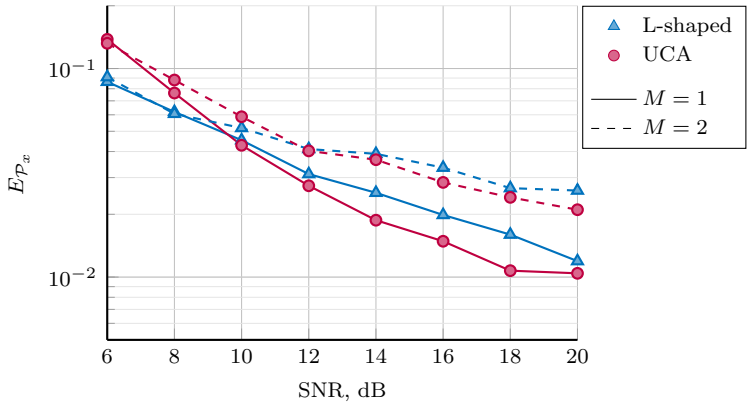


Figure 6.4: Angular-frequency spectrum reconstruction error vs. SNR for $M_{\mathrm{t}} = 14$.

consistent performance degradation for larger value of M, that increases with the increase of N. A possible explanation of this effect is that the noise at the different low-rate outputs of the same antenna originates from the same source and, hence, it is correlated. The noise at the outputs of different antennas on the other hand has different sources and therefore is likely to have independent statistics. This indicates that, other things being equal, the increase in the number of antennas is preferable to the increase of the number of sampling channels per antenna output.

6.3.4.3 Angular-frequency spectrum recovery

In the second round of simulations, we study the impact of the SNR on the performance of the 2D spectrum recovery measured by the MSE between the true and the estimated angular-frequency spectrum:

$$E_{P_x} = \frac{\frac{1}{N_\theta}\sum_{n=1}^{N_\theta} \|\mathcal{P}_x(\theta_n^{\mathrm{G}}, f) - \hat{\mathcal{P}}_x(\theta_n^{\mathrm{G}}, f)\|_2^2}{\|\mathcal{P}_x(\theta_n^{\mathrm{G}}, f)\|_2^2} \tag{6.29}$$

where $\theta_n^{\mathrm{G}} = \frac{\pi}{2N_\theta} n$. Figure 6.4 demonstrates the average MSE as a function of the SNR for $M \in [1, 2]$ and $M_t = 14$. It confirms that both considered types of antennas perform similarly in the considered scenario, while the increase of the antennas provides a slightly larger gain than the corresponding increase of the MWC channels per antenna.

6.4 Notes on possible extensions

6.4.1 A link to tensor-based processing

The multidimensional structure of the input signal $x(t)$ also provides opportunities to represent the observed data in a form of a tensor. Suppose for simplicity that $\mathbf{\Phi}_n = \mathbf{\Phi}$ for any $n = 1, 2, \ldots, N$, which is a sufficient condition for unique signal recovery as shown

by Theorem 7, and that $f_{\mathrm{c},k} \in \{pf_{\mathrm{p}}\}_{p=0}^{L_0}$. Then, we can write the DTFT $y_d(f)$ of the dth low-rate output as

$$y_d(f) = \sum_{k=1}^{K} c_{m,\ell_k} a_n(\theta_k, p_k f_{\mathrm{p}}) \mathcal{P}_x(\theta_k, f - \ell_k f_{\mathrm{p}}) = \boldsymbol{c}_m^{\mathrm{T}} \boldsymbol{A}_n \boldsymbol{x}(f), \tag{6.30}$$

where $p_k = f_{\mathrm{c},k}/f_{\mathrm{p}}$, $\boldsymbol{c}_m = [c_{m,\ell_1}, \ldots, c_{m,\ell_K}]^{\mathrm{T}}$, $\boldsymbol{A}_n$ is a $K \times K$ diagonal matrix that contains array responses $a_n(\theta_k, \ell_k f_{\mathrm{p}})$ on its main diagonal, while $\boldsymbol{x}(f)$ is a length-K vector with kth element $x_k(f) = X_k(f - f_{\mathrm{c},k})$. From M outputs of the nth antenna we obtain

$$\boldsymbol{y}_n(f) = \boldsymbol{C}\boldsymbol{A}_n \boldsymbol{x}(f) = \boldsymbol{C}\boldsymbol{X}(f)\boldsymbol{a}_n, \tag{6.31}$$

where $\boldsymbol{a}_n = \mathrm{diag}(\boldsymbol{A}_n)$ and $\boldsymbol{X}(f)$ is now a $K \times K$ diagonal matrix with $\boldsymbol{x}(f)$ on its main diagonal. The $M \times K$ matrix $\boldsymbol{C}$ in (6.31) is in turn comprised as $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M]^{\mathrm{T}}$.

To avoid possible confusion due the continuous frequency index f, let us represent the DTFT $\boldsymbol{y}_n(f)$ by F Fourier coefficients such that $\boldsymbol{y}_n(f)$ becomes a matrix $\boldsymbol{y}_n$ of size $M \times F$. THen, we can write it as follows

$$\boldsymbol{y}_n = \left[\boldsymbol{\mathcal{I}}_{3\times K} \times_1 \boldsymbol{C} \times_2 \boldsymbol{a}_n \times_3 \boldsymbol{X}\right]_{(3)}^{\mathrm{T}}, \tag{6.32}$$

where $\boldsymbol{X}$ is a $K \times F$ matrix representation of $\boldsymbol{x}(f)$ and $[\cdot]_{(n)}$ here denotes a mode-n tensor unfolding. Collecting all N outputs, we finally obtain

$$\boldsymbol{y} = \left[\boldsymbol{\mathcal{I}}_{3\times K} \times_1 \boldsymbol{C} \times_2 \boldsymbol{A} \times_3 \boldsymbol{X}\right]_{(3)}^{\mathrm{T}}, \tag{6.33}$$

with $\boldsymbol{A}$ being an $K \times N$ matrix of antenna responses $\boldsymbol{a}_n$ to K narrowband signals $x_k(t)$. Hence, if we could decompose $\boldsymbol{Y}$ as (6.33), it would allow us to find the parameters $\boldsymbol{\theta}$, $\boldsymbol{f}$ (that are encoded in $\boldsymbol{A}$) and the non-zero portions of $\mathcal{P}_x(\theta, f)$ which are contained in $\boldsymbol{X}$. From [309] it is known that the decomposition (6.33) is unique when $2K + 2 \leq \mathrm{krank}\,(\boldsymbol{C}) + \mathrm{krank}\,(\boldsymbol{A}) + \mathrm{krank}\,(\boldsymbol{X})$, which determines the interplay between the number of channels M, number of antennas N and conditions on $\boldsymbol{x}(f)$.

Furthermore, as the CS theory is being extended to include higher-order models [310]–[312], one could also think of writing (6.33) in an expanded form by representing $\boldsymbol{X}(f)$ as a sparse tensor, which would lead to a tensor-based sparse-recovery formulation.

6.4.2 3D sensing with polarimetric sources

So far, we have assumed that the antenna array is sensitive with respect to the azimuth angle of arrival θ only. While obtaining the sense of direction from which the signals arrive at the antenna is often most important with respect to the azimuthal antenna plane, the incoming waves are generally characterized not only by an azimuth angle of arrival and the central frequency, as we simplistically assumed in (6.3), but also by an elevation angle

of arrival as well as the polarization. It has been demonstrated that ignoring signals' and antenna's dependency on any of these parameters can have highly negative consequences for the estimation accuracy [313]. Having an adequate antenna description is especially important when dealing with practical antennas in realistic scenarios. The goal of this section, therefore, is to show how the full 3D polarimetric antenna description can be incorporated in the considered sensing framework.

6.4.2.1 Signal model extension

The array manifold is generally a function of the parameter vector $\boldsymbol{\gamma} = [\theta, \psi, f, \boldsymbol{p}^{\mathrm{T}}]^{\mathrm{T}}$ where, additionally to θ and f, we have the elevation angle of arrival ψ and the Jones vector $\boldsymbol{p} \in \mathbb{C}^{2\times 1}$ that describes the polarization state of the incident plain wave. Hence, a full description of the signal received at the nth antenna element in a sub-Nyquist multiband antenna system from Figure 6.1 is provided by

$$\begin{aligned} S_n(f) &= \sum_{k=1}^{K} a_n(\boldsymbol{\gamma_k}) \mathcal{P}_x(\theta_k, \psi_k, f) = \sum_{k=1}^{K} [a_n^{(\mathrm{V})}(\theta_k, \psi_k, f_{\mathrm{c},k}), a_n^{(\mathrm{H})}(\theta_k, \psi_k, f_{\mathrm{c},k})] \boldsymbol{p}_k \mathcal{P}_x(\theta_k, \psi_k, f) \\ &= \sum_{k=1}^{K} \boldsymbol{A}_n(\theta_k, \psi_k, f_{\mathrm{c},k}) \boldsymbol{p}_k \mathcal{P}_x(\theta_k, \psi_k, f), \end{aligned} \tag{6.34}$$

where $\boldsymbol{A}_n(\theta, \psi, f) = [a_n^{(\mathrm{V})}(\theta, \psi, f), a_n^{(\mathrm{H})}(\theta, \psi, f)] \in \mathbb{C}^{1\times 2}$ is the polarimetric response of the nth antenna element as a function of θ, ψ and f that consists of the antenna response for vertical $a_n^{(\mathrm{V})}(\theta, \psi, f)$ and horizontal $a_n^{(\mathrm{H})}(\theta, \psi, f)$ excitations. In contrast to (6.3), here the array manifold is a function of the full parameter vector $\boldsymbol{\gamma}$, where $\boldsymbol{a}(\boldsymbol{\gamma}) = \boldsymbol{a}(\theta, \psi, f, \boldsymbol{p})$, while the angular-frequency spectrum is a function of three variables[6], namely the two angles θ and ψ, and the frequency f, i.e.,

$$\mathcal{P}_x(\theta, \psi, f) = \sum_{k=1}^{K} X_k(f - f_{\mathrm{c},k}) \delta(\theta - \theta_k, \psi - \psi_k), \tag{6.35}$$

where $\delta(\theta, \psi)$ denotes a two-dimensional Kronecker delta function.

6.4.2.2 3D sensing in sub-Nyquist sampling framework

Consider now the receiver system in Figure 6.2. The relation between the angular-frequency spectrum $\mathcal{P}_x(\theta, \psi, f)$ of the multiband input $x(t)$ and the DTFTs of the antenna outputs $\boldsymbol{Y}(f)$ stays conceptually unchanged, i.e.,

$$\boldsymbol{y}(f) = \boldsymbol{\Psi} \boldsymbol{z}(f), \ f \in \mathcal{F}_{\mathrm{s}}. \qquad \text{(rep. of (6.10))}$$

[6] Note that we could include the polarization states of the incoming waves as a signal parameter as well, which would mean including the polarization domain into the sensing task. We assume here, however, that our sensing interests lie within angular and frequency domains only. Hence, we treat $\boldsymbol{p}_k$ as a nuisance parameter.

The elements of $\boldsymbol{\Psi}$ and $\boldsymbol{Z}(f)$, however, are now defined as

$$[\boldsymbol{\Psi}]_{d,g:g+1} = c_{d,\ell}\boldsymbol{A}_n(\theta_k, \psi_k, f_{\mathrm{c},k}), \text{ and} \tag{6.36}$$

$$z_{g:g+1}(f) = \boldsymbol{p}_k \mathcal{P}_x(\theta_k, \psi_k, f - \ell f_{\mathrm{p}}),\ f \in \mathcal{F}_{\mathrm{s}}, \tag{6.37}$$

respectively, where $d = (n-1)M + m$ and $g = (k-1)L + 2(L_0 + \ell) + 1$. Note that $\boldsymbol{\Psi}$ here us of size $NM \times 2KL$ while $\boldsymbol{z}(f)$ is now a vector of length $2KL$ with maximally $4K$ non-zeros that appear in blocks of 2 (one corresponding to the vertical and one to the horizontal signal component).

Given (6.36) and (6.37), both the dictionary $\boldsymbol{B}$ and the unknown vector $\boldsymbol{w}(f)$ in (6.13) have to be expanded to account for the presence of an additional search parameter ψ and the block structure of $\boldsymbol{z}(f)$. To do so, we introduce a grid of N_ϕ points in elevation with a grid step of $\psi_{\mathrm{d}} = 2\pi/N_\phi$ and build the $NM \times 2N_\theta N_\phi N_f L$ dictionary $\boldsymbol{B}$ such that its elements are given by

$$[\boldsymbol{B}]_{d,p:p+1} = c_{d,\ell}\boldsymbol{A}_n(\alpha_\theta \theta_{\mathrm{d}}, \alpha_\psi \psi_{\mathrm{d}}, \beta f_{\mathrm{d}}), \tag{6.38}$$

where $p = (\alpha\phi - 1)N_\phi N_F L + (\alpha_\psi - 1)N_F L + \beta L + 2(L_0 + \ell) + 1$, whereas $\boldsymbol{w}(f)$ accordingly expands into an $2N_\theta N_\phi N_f L \times 1$ vector that is $2K$ 2-block sparse. Note that the conditions of Theorems 7 and 8 hold in this case as well. The recovery approaches described in Section 6.3 are also directly applicable here, but with a recommended change to the block-sparse adaptations of the sparse recovery algorithms that are used to estimate the support of $\boldsymbol{w}(f)$, e.g., by applying block-OMP (BOMP) [47] for instance.

It is important to note that adding a second angular dimension increases the search space for angular-frequency sensing even further. This highlights the importance of the development of alternative recovery methods, such as the ones that build on the ideas of tensor-based processing mentioned above, which could allow to avoid comprising large two-dimensional dictionaries for the recovery of signals with multi-dimensional structure.

6.5 Concluding remarks

In this chapter, we discussed the task of joint angular-spectral sensing in a multiband setting. In order to obtain angular information about the incoming waves, we employed a (wideband) antenna array and developed a system based on the MWC sub-Nyquist sampling architecture for 2D angular-frequency sensing from sub-Nyquist samples. We then described a generic receiver structure, presented a frequency domain analysis of its operation and derived conditions on the total number of sampling channels (which in our system is defined by the number of antennas times the number of MWC sampling channels per antenna) for perfect 2D spectrum recovery in a noiseless environment. Following this analysis, we suggested two particular methods to estimate the central frequencies and the DoAs of the individual signals occupying the active sub-bands. In the first one, we formulate the parameter estimation task

as a sparse recovery problem where the search dictionary is comprised of system responses to all possible combinations of parameter values. In the second method, we first estimate the frequency support of a multiband input from the outputs of a single antenna (as in the regular MWC) and then solve the associated parameter estimation problem for the identified active spectral cells only. Additionally, we outlined possible alternative ways to perform parameter estimation in the proposed system, which build upon the ideas from tensor-based processing and could help in avoiding the need for a discrete estimation grid and a large parametric dictionary. Importantly, none of the discussed methods assumes a specific structure of the antenna array responses, such as uniform sensor spacing or isotropic radiation patterns of the individual elements, as in the earlier works discussed at the beginning of this chapter. In contrast, they can be applied to (almost) arbitrary antenna arrays. Although in this thesis we primarily concentrated on non-polarimetric (multiband) azimuth angle of arrival estimation, the described system allows the inclusion of full 3D polarimetric array response to reflect the sensitivity of antennas with respect to both azimuth and elevation angles of arrival, as well as polarization states of the incoming waves, which is known to have a significant impact on the accuracy of the DoA estimation [313]. This said, it also potentially allows the incorporation of measured antennas, using the EADF approach from [314] for instance.

By performing the analysis presented in this chapter, we take first steps towards a more comprehensive understanding of a sub-Nyquist multiband antenna system for angular-spectral sensing. While the proposed system offers a certain degree of design flexibility (in terms of the number of antenna elements, the number of MWC channels and the choice of the mixing sequences) and can potentially encompass different array types, there is a need for better understanding of its performance and limitations, especially with respect to the characteristics of real antenna arrays. Note that the main goal of the considered system is to be able to observe a large frequency band that could potentially span up to several gigahertz. Design of an antenna for such a wide frequency range is a research task by itself; how well its characteristics could conform to the considered multiband scenario remains an open question.

Chapter 7

Multiband spectrum sensing and time delay estimation

7.1 Motivation and related work

If equipping a mobile radio sensor with a *sense of direction* can be advantageous in practical scenarios, providing it with an awareness of the spatial positions of the transmitting sources offers even greater opportunities [315]. Location awareness can be obtained from a database, in case of stationary sources with known geo-locations, or by means of radio localization. In view of the latter, methods based on time estimation are often considered particularly suitable, especially for applications such as the cognitive radio (CR) [316], [317]. While there is ample literature on emitter localization in various scenarios and a large body of research is dedicated to wideband multiband spectrum sensing, the literature on multiband time delay based emitter localization is comparatively scarce. It seems that the only exception is the so-called localization over dispersed (or concatenated) spectrum where the emitter to be localized is assumed to be simultaneously transmitting over multiple non-contiguous frequency bands [318]–[320]. Such a multiband operation creates a larger virtual bandwidth and facilitates propagation diversity, which is used for resolution enhancement. Another somewhat-related work [321] considers multiple wideband source localization from frequency domain components where each source occupies the entire observed frequency range.

In a multiband sensing considered in this thesis, our goal is different. We aim at estimating time delays (and the central frequencies) of a superposition of multiple unknown transmissions that originate from different sources and occupy distinct narrowband communication sub-bands. In order to obtain time delay information, we employ a network of spatially distributed time-synchronized sensing nodes that acquire and process the wideband multiband input at a sub-Nyquist rate. As before, the sensors utilize the MWC architecture from [255] for sub-Nyquist sampling. Capitalizing on the results presented in [256], [257], where it is shown that the power spectrum of a wideband signal can be efficiently reconstructed

from low-rate cross-correlations between different outputs of a single sub-Nyquist sensor, we use outputs of (different) spatially distributed sub-Nyquist sensing nodes to estimate the central frequencies and the relative autocorrelation functions of the individual signals within the active communication sub-bands. Particularly, we propose two ways of approaching this task. In the first one, we jointly recover the frequency support and the corresponding low-rate relative autocorrelation functions from the cross-correlations between outputs of different sensors. In the second method, we first estimate the frequency support, recover the respective time signals from the outputs of each sensor, and then compute the low-rate relative autocorrelation functions for each of the signals independently. Once the relative autocorrelation functions are found, in the line of sight (LoS) we can determine the time difference of arrivals (TDoAs) within each sub-band, e.g., by detecting the time lag of the largest peak of the relative autocorrelation function, and proceed with emitter localization using available approaches, as discussed in [322] for instance. Otherwise, one could potentially use the (partial) knowledge of the channel impulse response structure (obtained from channel measurements, geometry-based modelling and/or ray-tracing tools) to estimate emitters' locations from the obtained relative autocorrelation functions, e.g., using a multilateration method as in [323].

For accurate time delay based emitter localization in a distributed sensing network, it is desirable to ensure that for localization purposes we only use relative autocorrelation functions that describe the same sets of emitters. Note that some communication sub-bands can be occupied by different transmitters at different locations while not causing local in-channel interference due to the propagation conditions, e.g., when two terminals are blocked from each other by an obstacle or separated by a large enough distance. When this goes undetected, a situation can occur when two sensors observe two (partially) different sets of transmitters that appear to be the same. This can potentially result in erroneous estimates, appearance of ghost paths and overall performance degradation. Generally, detecting such in-band interference is not an easy task since both transmitters operate in the same frequency band, possibly using a similar transmission scheme. An availability of DoA estimation capabilities at the sensors could potentially help to resolve the two sources in this situation. It turns out that the sub-Nyquist sampling framework also offers opportunity to detect the presence of in-band interference, provided that the number of interfering sources is low. In this chapter, we propose a method that exploits the structure of the covariance matrices computed from the outputs of different sensors to detect the presence of possible in-band interference and estimate the central frequencies of the sub-bands it occurs in.

The material of this chapter is partially presented in [A10] and [A8], while [A7] considers the task of hybrid TDoA/DoA localization from compressed measurements.

7.2 Multiband time delay estimation from sub-Nyquist samples

7.2.1 Preliminaries

Consider a sensing system of P spatially distributed wideband sensing nodes that simultaneously observe a wide frequency band organized into multiple communication sub-bands, as schematically illustrated in Figure 7.1. The nodes are time synchronized by a common clock source and are able to exchange measurement data with each other or some centralized processing station. The goal of the sensing system is to detect which sub-bands are in use and estimate the relative time delays of the transmissions that occupy them from sub-Nyquist samples of the wideband multiband input. With this in mind, we begin our study with re-visiting the multiband signal model in view of the time delay estimation task and introducing the basics of time-delay estimation in a distributed sensor network.

7.2.1.1 Input signal

Consider wide frequency band $\mathcal{F} = [0, W]$ that is organized into N consecutive (narrowband) communication sub-bands of bandwidth W/N and central frequencies $f_n^{\rm c}, n = 1, 2, \ldots, N$. Out of N sub-channels, K are occupied by independent uncorrelated complex-valued transmissions $x_k(t)$ with distinct central frequencies[1] $f_{{\rm c},k} \in \{f_n^{\rm c}\}_{n=1}^N$. Each transmission $x_k(t)$ is assumed to be zero-mean wide-sense stationary with a bandwidth not exceeding $B \leq W/N$, while the central frequency of the sub-channel it occupies and its exact baseband waveform $\bar{x}(t)$ are

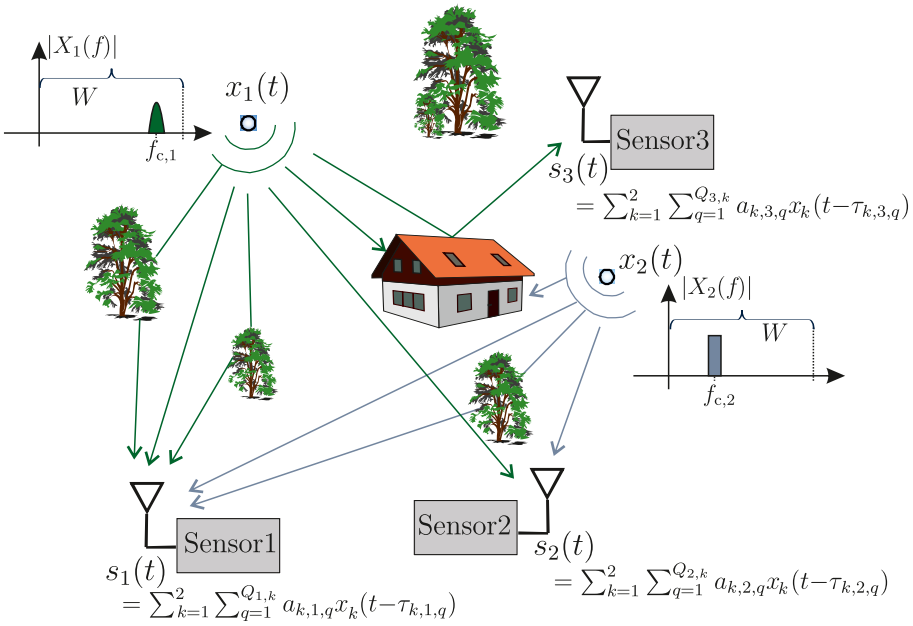


Figure 7.1: An example of a distributed sensing system with $P = 3$ sensing nodes observing $K = 2$ transmissions with unknown central frequencies in $\mathcal{F} = [0, W]$.

[1] Note that here we consider the case when the individual signals occupy exactly one spectral cell.

assumed unknown such that the Fourier transform $S(f)$ of the total wideband input $s(t)$, given by

$$s(t) = \sum_{k=1}^{K} x_k(t) = \sum_{k=1}^{K} \bar{x}_k(t) e^{2\pi \jmath f_{\mathrm{c},k} \tau}, \tag{7.1}$$

is identically zero anywhere outside $\mathcal{F}$. For each $x_k(t)$ we define an autocorrelation function $R_{kk}(\tau)$ as

$$R_{kk}(\tau) \triangleq \mathbb{E}\{x_k(t) x_k^*(t-\tau)\} = \mathbb{E}\{\bar{x}_k(t) \bar{x}_k^*(t-\tau) e^{2\pi \jmath f_{\mathrm{c},k}(t-t+\tau)}\} = \bar{R}_{kk}(\tau) e^{2\pi \jmath f_{\mathrm{c},k} \tau}, \tag{7.2}$$

where $\bar{R}_{kk}(\tau) = \mathbb{E}\{\bar{x}_k(t) \bar{x}_k^*(t-\tau)\}$. The Fourier transform of $R_{kk}(\tau)$,

$$P_{kk}(f) = \int_{-\infty}^{\infty} R_{kk}(\tau) e^{-2\pi \jmath f \tau} \mathsf{d}\tau = \int_{-\infty}^{\infty} \bar{R}_{kk}(\tau) e^{-2\pi \jmath (f - f_{\mathrm{c},k}) \tau} \mathsf{d}\tau = \bar{P}_{kk}(f - f_{\mathrm{c},k}), \tag{7.3}$$

is zero for all $f \notin \mathcal{F}$, whereas its support is confined within an interval of length B. Finally, we note that since $x_k(t)$ are independent and uncorrelated it holds that

$$\forall\, k_1, k_2 \in [1, K] :\ k_1 \neq k_2\ R_{k_1 k_2}(\tau) = \mathbb{E}\{x_{k_1}(t) x_{k_2}^*(t-\tau)\} \equiv 0. \tag{7.4}$$

7.2.1.2 Received signal

Let $h_{k,p}(t)$ denote an impulse response of the propagation channel between a pth sensing node and a kth signal source and $Q_{k,p} \leq Q$ be the number of multipath components it contains. Assuming for simplicity time-invariant channels such that each multipath component is represented by two constants, its amplitude $a_{k,p,q}$ and delay $\tau_{k,p,q}$, the noise-free[2] signal received by the pth sensor can be written as

$$s_p(t) = \sum_{k=1}^{K} x_k(t) * h_{k,p}(t) = \sum_{k=1}^{K} \sum_{q=1}^{Q_{p,k}} a_{k,p,q} x_k(t - \tau_{k,p,q}) = \sum_{k=1}^{K} x_{k,p}(t), \tag{7.5}$$

where $x_{k,p}(t) = \sum_{q=1}^{Q_{p,k}} a_{k,p,q} x_k(t - \tau_{k,p,q})$. Note that the delay $\tau_{k,p,q}$ is assumed to be growing with index q meaning that $\tau_{k,p,1} < \tau_{k,p,2} < \cdots < \tau_{k,p,Q_{k,p}}$. Given (7.5), we can calculate the autocorrelation function of $s_p(t)$ as

$$\begin{aligned} R_{pp}(\tau) &= \mathbb{E}\{s_p(t) s_p^*(t-\tau)\} = \sum_{k=1}^{K} \mathbb{E}\{x_{k,p}(t) x_{k,p}^*(t-\tau)\} \\ &= \sum_{k=1}^{K} \sum_{q_1, q_2} a_{k,p,q_1} a_{k,p,q_2}^* R_{kk}(\tau - \tau_{k,p,q_1} + \tau_{k,p,q_2}) = \sum_{k=1}^{K} \sum_{q_1, q_2} \tilde{a}(p)_{k,q_1,q_2} R_{kk}(\tau - \tilde{\tau}_{k,q_1,q_2}^{(p)}), \end{aligned} \tag{7.6}$$

[2]Note that practically, the signal $s_p(t)$ would be contaminated with noise, while the noise free setting is adopted here for the sake of clarity only.

where $\tilde{a}(p)_{k,q_1,q_2} = a_{k,p,q_1} a^*_{k,p,q_2}$ and $\tilde{\tau}^{(p)}_{k,q_1,q_2} = \tau_{k,p,q_1} - \tau_{k,p,q_2}$ is the relative delay between different signal copies at pth sensor, while $q_1, q_2 \in [1, Q_{k,p}]$. Note that due to (7.4) the cross-correlation terms between delayed copies of different input signals $x_k(t)$ are zero, whereas the local extrema of $|R_{pp}(\tau)|$ occur at time instances $\tilde{\tau}^{(p)}_{k,q_1,q_2}$ symmetrically with respect to the origin.

7.2.1.3 Relative time delay estimation

For any pair of sensors (p_1, p_2), we can also define a cross-correlation function $R_{p_1,p_2}(t)$ as

$$\begin{aligned} R_{p_1,p_2}(\tau) &= \mathbb{E}\{s_{p_1}(t) s^*_{p_2}(t-\tau)\} = \sum_{k=1}^{K} \mathbb{E}\{x_{k,p_1}(t) x^*_{k,p_1}(t-\tau)\} \\ &= \sum_{k=1}^{K} \sum_{q_1,q_2} a_{k,p_1,q_1} a^*_{k,p_2,q_2} R_{kk}(\tau - \tau_{k,p_1,q_1} + \tau_{k,p_2,q_2}) \\ &= \sum_{k=1}^{K} \sum_{q_1,q_2} \tilde{a}^{(p_1,p_2)}_{k,q_1,q_2} R_{kk}\left(\tau - \tilde{\tau}^{(p_1,p_2)}_{k,q_1,q_2}\right) = \sum_{k=1}^{K} R^{(p_1,p_2)}_k(\tau). \end{aligned} \tag{7.7}$$

Here, $R^{(p_1,p_2)}_k(\tau)$ denotes the relative autocorrelation function of the kth input signal $x_k(t)$ for sensors p_1, p_2, i.e.,

$$\begin{aligned} R^{(p_1,p_2)}_k(\tau) &= \sum_{q_1,q_2} \tilde{a}^{(p_1,p_2)}_{k,q_1,q_2} R_{kk}\left(\tau - \tilde{\tau}^{(p_1,p_2)}_{k,q_1,q_2}\right) \\ &= \sum_{q_1,q_2} \tilde{a}^{(p_1,p_2)}_{k,q_1,q_2} \bar{R}_{kk}\left(\tau - \tilde{\tau}^{(p_1,p_2)}_{k,q_1,q_2}\right) e^{2\pi \jmath f_{\mathrm{c},k}\tau} = \bar{R}^{(p_1,p_2)}_k(\tau) e^{2\pi \jmath f_{\mathrm{c},k}\tau}, \end{aligned} \tag{7.8}$$

while $\tilde{a}^{(p_1,p_2)}_{k,q_1,q_2} = a_{k,p_1,q_1} a^*_{k,p_2,q_2}$ and $\tilde{\tau}^{(p_1,p_2)}_{k,q_1,q_2} = \tau_{k,p_1,q_1} - \tau_{k,p_2,q_2}$ is the relative delay between different signal copies at two different sensors with indices p_1 and p_2.

In the presence of direct paths to all K signal sources, the relative time delays $\tilde{\tau}^{(p_1,p_2)}_{k,1,1}$ are commonly referred to as TDoAs [324]. In LoS conditions, they can be used for emitter localization, provided that the sensors have knowledge of their relative positions [322]. In case of non line of sight (NLoS) on the other hand, one could potentially use (partial) knowledge of the channel impulse response structure (obtained from channel measurements, geometry-based modelling and/or ray-tracing tools) to estimate emitters' locations from the peaks in $\bar{R}^{(p_1,p_2)}_k(\tau)$. This being said, availability of the relative autocorrelation functions $R^{(p_1,p_2)}_k(\tau)$ is prerequisite for time delay based emitter localization. From (7.8), $R^{(p_1,p_2)}_k(\tau)$ is fully defined by its baseband version $\bar{R}^{(p_1,p_2)}_k(\tau)$ and the central frequency $f_{\mathrm{c},k}$. Our goal is therefore to jointly estimate the central frequencies $f_{\mathrm{c},k}$ and the baseband relative autocorrelation functions $\bar{R}^{(p_1,p_2)}_k(\tau)$ of the signal sources from sub-Nyquist samples of $s_p(t)$.

7.2.2 Multiband time delay estimation

7.2.2.1 Sub-Nyquist receiver system

To acquire the wideband multiband input, each sensor employs the MWC sub-Nyquist sampling architecture [255]. Since we have discussed the MWC in details in previous chapters, we do not repeat the description of its operation here once again but rather refer the reader to Section 5.3.1. Note, however, that in the following we assume that the number of sampling channels M, the sampling sequences $p_m(t)$ and the sampling rate f_s are the same[3] for all sensors, while $f_\mathrm{s} = f_\mathrm{p} \approx B$.

Frequency-domain description. Consider the mth MWC sampling channel of the pth sensing node and let $y_{p,m}[t_n = nT_\mathrm{s}]$ denote its (discrete) output, where $T_\mathrm{s} = 1/f_\mathrm{s}$ is the sampling interval. We know by now that since each $p_m(t)$ is T_p-periodic, it can be represented as

$$p_m(t) = \sum_{\ell=-\infty}^{\infty} c_{m,\ell} e^{\jmath 2\pi \ell f_\mathrm{p} t} \text{ where } c_{m,\ell} = \frac{1}{T_\mathrm{p}} \int_0^{T_\mathrm{p}} p_m(t) e^{-\jmath 2\pi \ell f_\mathrm{p} t} \mathsf{d}t. \tag{7.9}$$

Repeating the analysis from Section 5.3.1, we get the following relation between the (known) DTFT of the output samples $y_{p,m}[t_n]$ and the (unknown) spectrum of $s_p(t)$

$$y_{p,m}(f) = \boldsymbol{c}_m^\mathrm{H} \boldsymbol{z}_p(f),\ f \in \mathcal{F}_\mathrm{s} = [-f_\mathrm{s}/2, f_\mathrm{s}/2]. \tag{7.10}$$

In line with previous notations, $y_{p,m}(f)$ here represents the DTFT of $y_{p,m}[t_n]$ and $\boldsymbol{c}_m$ is a length-L vector that contains the Fourier coefficients $c_{m,\ell}$. The unknown vector $\boldsymbol{z}_p(f)$ contains f_p-shifted low-passed filtered copies of $s_p(f)$ such that

$$z_{p,\ell}(f) = s_p(f - \ell f_\mathrm{p}),\ f \in \mathcal{F}_s. \tag{7.11}$$

Time-domain description. Writing (7.10) we obtain

$$y_{p,m}[t_n] = \boldsymbol{c}_m^H \boldsymbol{z}_p[t_n], \tag{7.12}$$

where $\boldsymbol{z}_p[t_n]$ is a length-L vector with ℓth element being the IDTFT of $z_{p,\ell}(f)$. Note that since $f_{\mathrm{c},k} \in \{f_n^\mathrm{c}\}_{n=1}^N$ and $f_\mathrm{p} = f_\mathrm{s} \approx B$, only K entries of $\boldsymbol{z}_p[t_n]$ are non-zero, namely the ones with indices in the set $\mathcal{S} = \{\ell_k = f_{\mathrm{c},k}/f_\mathrm{p}\}_{k=1}^K$. Furthermore, according to (7.11), the ℓ_kth entry $z_{p,\ell_k}[t_n]$ contains the $(\ell_k f_\mathrm{p})$-shifted low-pass filtered copy of $s_p(t)$. Hence, (7.12) can

[3]Note that the first two assumptions are not formal requirements for the operation of the proposed time delay estimation system but rather adopted for simplicity of presentation. Both the number of sampling channels and the mixing sequences can differ from sensor to sensor. Conceptually, the sampling rate can also be different for different sensors, provided that it is an odd multiple of f_p, which is kept the same for all sensors.

be equivalently represented as

$$y_{p,m}[t_n] = \sum_{k=1}^{K} c_{m,\ell_k} z_{p,\ell_k}[t_n] = \sum_{k=1}^{K} c_{m,\ell_k} \bar{x}_{k,p}[t_n] = \sum_{k=1}^{K} c_{m,\ell_k} \sum_{q=1}^{Q_{p,k}} a_{k,p,q} \bar{x}_k[t_n - \tau_{k,p,q}], \quad (7.13)$$

where $\bar{x}_{k,p}[t_n] = \bar{x}_{k,p}[nT_{\rm s}]$ and $\bar{x}_k[t_n] = \bar{x}_k[nT_{\rm s}]$ are the baseband (low-rate) sampled equivalents of $x_{k,p}(t)$ and $x_k(t)$, respectively.

7.2.2.2 Central frequency and autocorrelation estimation

Below, we propose two methods to estimate $f_{{\rm c},k}$ and $\bar{R}_k^{(p_1,p_2)}(\tau)$ from $\{y_{p,m}[t_n]\}_{p,m}$. In the first one, we recover the frequency support and the relative autocorrelation functions jointly, while in the second method we adopt a two-step approach and estimate them separately.

Joint recovery. Consider correlation between the ith and jth outputs of sensors p_1, p_2:

$$\begin{aligned} r_{i,j}^{(p_1,p_2)}[\nu T_{\rm s}] &= \mathbb{E}\{y_{p_1,i}[t_n] y_{p_2,j}^*[t_n - \nu T_{\rm s}]\} = \sum_{k=1}^{K} w_{i,j,\ell_k} \sum_{q_1,q_2} \tilde{a}_{k,q_1,q_2}^{(p_1,p_2)} \mathbb{E}\{\bar{x}_k[t_n] \bar{x}_k^*[t_n - \nu T_{\rm s} - \tilde{\tau}_{k,q_1,q_2}^{(p_1,p_2)}]\} \\ &= \sum_{k=1}^{K} w_{i,j,\ell_k} \sum_{q_1,q_2} \tilde{a}_{k,q_1,q_2}^{(p_1,p_2)} \bar{r}_{kk}[\nu T_{\rm s} - \tilde{\tau}_{n,q_1,q_2}^{p_1,p_2}] = \sum_{k=1}^{K} w_{i,j,\ell_k} \bar{r}_k^{(p_1,p_2)}[\nu T_{\rm s}], \end{aligned} \quad (7.14)$$

where $w_{i,j,\ell_k} = c_{i,\ell_k} c_{j,\ell_k}^*$. Here, $\bar{r}_{kk}[\nu T_{\rm s}]$ represents the autocorrelation function of $\bar{x}_k[t_n]$, which, in turn, corresponds to the low-rate sampled baseband version of $R_{kk}(\tau)$. Similarly, $\bar{r}_k^{(p_1,p_2)}[\nu T_{\rm s}] = \sum_{q_1,q_2} \tilde{a}_{k,q_1,q_2}^{p_1,p_2} \bar{r}_{kk}[\nu T_{\rm s} - \tilde{\tau}_{k,q_1,q_2}^{(p_1,p_2)}]$ is the low-rate baseband version of $R_k^{(p_1,p_2)}(\tau)$. Now, since $f_{\rm s} \geq B$ to determine $R_k^{(p_1,p_2)}(\tau)$ it suffices to estimate $\bar{r}_k^{(p_1,p_2)}[\nu T_{\rm s}]$.

Let $\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[nT_{\rm s}]$ denote an $L \times 1$ vector whose (ℓ_k)th element is $\bar{r}_k^{(p_1,p_2)}[\nu T_{\rm s}]$, while the rest are identically zero, i.e.,

$$r_{\mathbf{z},\ell}^{(p_1,p_2)}[nT_{\rm s}] = \begin{cases} \bar{r}_k^{(p_1,p_2)}[\nu T_{\rm s}], & \ell = \ell_k \\ 0, & \text{otherwise} \end{cases}. \quad (7.15)$$

Stacking different cross-correlations $r_{i,j}^{(p_1,p_2)}[\nu T_{\rm s}]$, $i,j = 1,2\ldots,M$ together into one vector $\boldsymbol{r}^{(p_1,p_2)}[nT_{\rm s}]$ of length M^2 we obtain

$$\boldsymbol{r}_{\mathbf{y}}^{(p_1,p_2)}[\nu T_{\rm s}] = \boldsymbol{W} \boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\rm s}]. \quad (7.16)$$

The $M^2 \times L$ matrix $\boldsymbol{W}$ in (7.16) is comprised of elements $w_{i,j,\ell}$ in such a way that its ℓth column $\boldsymbol{w}_\ell$ is given by

$$\boldsymbol{w}_\ell = \Big[\underbrace{w_{1,1,\ell},\ldots,w_{1,M,\ell}}_{M}, \underbrace{w_{2,1,\ell},\ldots,w_{2,M,\ell}}_{M}, \ldots, \underbrace{w_{M,1,\ell},\ldots,w_{M,M,\ell}}_{M}\Big]^{\rm T}. \quad (7.17)$$

We notice that the unknown vector $\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ is K-sparse as it contains non-zero relative autocorrelation functions $\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ at the positions corresponding to the frequency cells with indices ℓ_k. Therefore, we can use standard CS techniques to recover the sparse vector $\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ for each ν independently. For instance, we could solve the following problem

$$\hat{\boldsymbol{r}}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}] = \arg \min_{\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]} \left(\|\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]\|_1 \right) \quad \text{s.t} \quad \boldsymbol{r}_{\mathbf{y}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}] = \boldsymbol{W} \boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]. \quad (7.18)$$

Alternatively, we can take advantage of the joint sparse structure of $\boldsymbol{r}_{\mathbf{z}}^{p_1,p_2}[nT_{\mathrm{s}}]$ and apply the CTF block to convert (7.16) into an MMV problem, as we did for spectral recovery and joint angular-spectral recovery in Sections 5.3 and 6.3, respectively. With (7.16) in mind, the support recovery step in this case amounts to finding the (row) sparsest solution to

$$\boldsymbol{V} = \boldsymbol{W}\boldsymbol{U}, \text{ where } \boldsymbol{V} \text{ satisfies} \quad (7.19)$$

$$\begin{aligned} \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} &= \sum_{\nu=-\infty}^{\infty} \boldsymbol{r}_{\mathbf{y}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}](\boldsymbol{r}_{\mathbf{y}}^{(p_1,p_2)})^{\mathrm{T}}[\nu T_{\mathrm{s}}] = \sum_{\nu=-\infty}^{\infty} \boldsymbol{W}\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}](\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)})^{\mathrm{T}}[\nu T_{\mathrm{s}}]\boldsymbol{W}^{\mathrm{H}} \\ &= \boldsymbol{W}\boldsymbol{Q}_{\mathbf{z}}^{(p_1,p_2)}\boldsymbol{W}^{\mathrm{H}}, \end{aligned} \quad (7.20)$$

while $\boldsymbol{Q}_{\mathbf{z}}^{(p_1p_2)} = \sum_{\nu} \boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}](\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)})^{\mathrm{T}}[\nu T_{\mathrm{s}}] \in \mathbb{C}^{L\times L}$. We know that the support of the unique sparse solution to (7.19) coincides with the support $\mathcal{S} = \{\ell_k\}_{k=1}^K$ of $\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$. Note that $\mathcal{S}$ contains the indices of the central frequencies f_k of the K active central frequencies $f_{\mathrm{c},k} = \ell_k f_{\mathrm{p}}$, which means that by identifying the support of $\boldsymbol{r}_{\mathbf{z}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ we determine the spectral support of the wideband input $s(t)$.

After $\mathcal{S}$ is found, we can obtain the individual cross-correlation functions $\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ by inverting $\boldsymbol{W}_{\mathcal{S}}$ as

$$\left[\bar{r}_{\mathbf{z}}^{(p_1,p_2)}\right]_{\mathcal{S}}[\nu T_{\mathrm{s}}] = \boldsymbol{W}_{\mathcal{S}}^{\dagger}\boldsymbol{r}_{\mathbf{y}}^{(p_1,p_2)}[\nu T_{\mathrm{s}}], \quad (7.21)$$

where $(\cdot)^{\dagger}$ stands for the (Moore-Penrose) matrix pseudo-inverse, whereas $\left[\bar{r}_{\mathbf{z}}^{(p_1,p_2)}\right]_{\mathcal{S}}$ and $\boldsymbol{W}_{\mathcal{S}}$ denote the entries of $\bar{r}_{\mathbf{z}}^{(p_1,p_2)}$ and the columns of $\boldsymbol{W}$ indexed by $\mathcal{S}$, respectively. Note that once $\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ are known we can estimate the TDoAs $\tilde{\tau}_{k,1,1}^{(p_1,p_2)}$ by any of the methods available for identifying direct-path time-delays in multipath, such as the one proposed in [324] for instance.

Two-step estimation Here we propose another way of approaching the problem of estimating $f_{\mathrm{c},k}$ and $\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ from the low-rate outputs $y_{p,m}[t_n]$. It is based on the following two-step procedure: i) we first estimate the support $\mathcal{S}$ and the corresponding low-rate sequences $\bar{x}_{k,p}[t_n]$ from the output of each individual sensor, and then ii) compute the relative autocorrelation functions $\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ for each k independently.

We begin by collecting all M outputs $y_{p,m}[t_n]$ of the pth receiver together into one length-M vector $\boldsymbol{y}_m[t_n]$:

$$\boldsymbol{y}_p[t_n] = \boldsymbol{\Phi}\boldsymbol{z}_p[t_n], \tag{7.22}$$

where $\boldsymbol{\Phi} = [\boldsymbol{c}_1, \cdots, \boldsymbol{c}_M]^{\mathrm{H}} \in \mathbb{C}^{M\times L}$. Obviously, the support of $\boldsymbol{z}_p[t_n]$ is given by $\mathcal{S}$, since $z_{p,\ell}[t_n]$ is the IDTFT of $z_{p,\ell}(f)$ defined by (7.11).

Similarly to (7.16), one can solve (7.22) either for each t_n independently or by applying the CTF block as in (7.20) where instead of $\boldsymbol{W}$ we have $\boldsymbol{\Phi}$ and $\boldsymbol{V}$ now satisfies

$$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \sum_{t_n=-\infty}^{\infty} \boldsymbol{y}_p[t_n]\boldsymbol{y}_p^{\mathrm{H}}[t_n] = \boldsymbol{\Phi}\boldsymbol{Q}_{\boldsymbol{z}}^{(p)}\boldsymbol{\Phi}^{\mathrm{H}} \tag{7.23}$$

with $\boldsymbol{Q}_{\boldsymbol{z}}^{(p)} = \sum_{t_n} \boldsymbol{z}_p[t_n]\boldsymbol{z}_p^{\mathrm{H}}[t_n] \in \mathbb{C}^{L\times L}$. Once we find the support $\mathcal{S}$, we can estimate the individual low-rate sequences $\bar{x}_{k,p}[t_n]$ via

$$[\boldsymbol{z}_p]_{\mathcal{S}}[t_n] = \boldsymbol{\Phi}_{\mathcal{S}}^{\dagger}\boldsymbol{y}_p[t_n], \tag{7.24}$$

and then obtain the low-rate relative autocorrelations $\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ as

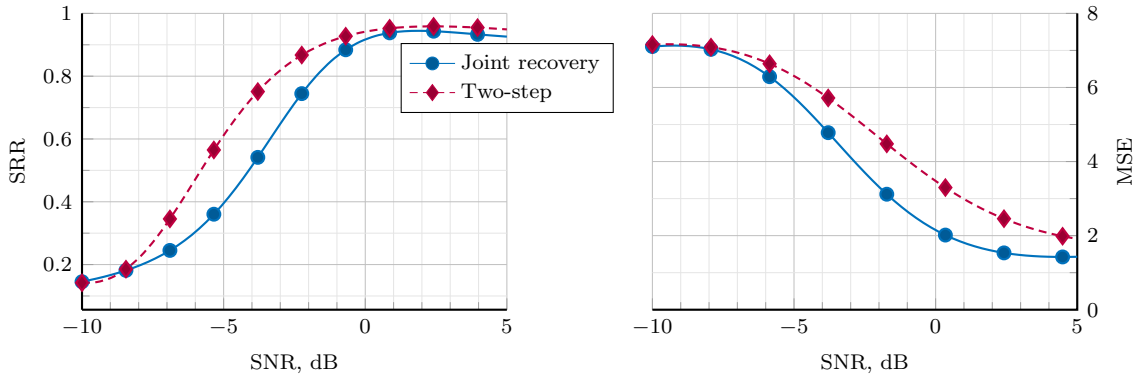
$$\bar{r}_k^{(p_1,p_2)}[\nu T_{\mathrm{s}}] = \mathbb{E}\{\bar{x}_{k,p_1}[t_n]\bar{x}_{k,p_2}[t_n - \nu T_{\mathrm{s}}]\}. \tag{7.25}$$

Note that the minimal number of measurements for the two-step approach increases to $M > 2K$, compared to $M > \sqrt{2K}$ in the joint recovery approach discussed above. This is due to the fact that here we only use M outputs of the same sensor to estimate the frequency support. On the other hand, if the support in (7.23) is estimated correctly, using (7.25) may prove advantageous under a strict limitation on the sensing time[4].

7.2.3 Performance evaluation

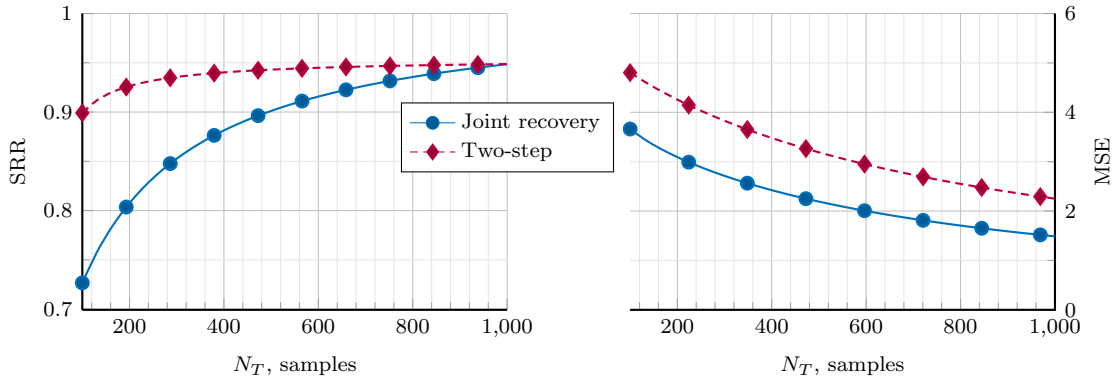
Simulation setup For numerical evaluation we consider a frequency band of $W = 3.9$ GHz that is split into $N = 135$ communication channels. The band W is occupied by $K = 3$ BPSK modulated signals $x_k(t)$, each with a bandwidth of $B = 20$ MHz and a distinct carrier $f_{\mathrm{c},k}$ chosen uniformly at random from $\{f_n^{\mathrm{c}}\}_{n=1}^N$. The sensors operate with $M = 20$ sampling channels at the sampling rate of $f_{\mathrm{s}} = 28$ MHz each, while the mixing sequences $p_m(t)$ are generated as pseudo-random $\{\pm 1\}$ piece-wise constant functions with a period of $T_{\mathrm{p}} = T_{\mathrm{s}}$. The total sampling rate at each sensor is 560 MHz, which is 14% of the Nyquist rate. The number of multipath components is assumed to be $Q_{k,p} = 2$ for all propagation channels, while the corresponding time delays $\tau_{k,p,q}$ and amplitudes $a_{k,p,q}$ are chosen uniformly at

[4]Note that in practice, the cross-correlation functions $r_{i,j}^{(p_1,p_2)}[\nu T_{\mathrm{s}}]$ in (7.14) are estimated from a limited number of samples (over a limited sensing time), which potentially creates non-zero cross-correlation terms between the delayed copies of different input signals $\bar{x}_k[t_n]$. This is not a problem for the two-step approach since the relative autocorrelation functions there are estimated separately for each sub-band.



(a) SRR **(b)** MSE

Figure 7.2: Average SRR (a) and normalized MSE (b) as a function of SNR.



(a) SRR **(b)** MSE

Figure 7.3: Average SRR (a) and normalized MSE (b) as a function of N_{T} for an SNR of 0 dB.

random from $[\frac{N_{\mathrm{T}}}{100}, \frac{9N_{\mathrm{T}}}{100}]$ and $[0.6, 1]$, where N_{T} denotes the sensing time in samples.

Due to the finite sensing time, the cross-correlation functions (7.14) are estimated as

$$\mathrm{MSE} = \hat{r}_{i,j}^{(p_1,p_2)}[\tau_\nu] = \frac{1}{N_{\mathrm{T}}} \sum_{\varsigma=1}^{N_{\mathrm{T}}} y_{p_1,i}[t_\varsigma] y_{p_2,j}[t_\varsigma - \tau_\nu], \tag{7.26}$$

while (7.20) and (7.23) are also computed over finite windows $\nu \in [1, N_{\mathrm{T}}]$ and $t_i \in [t_0, t_{N_{\mathrm{T}}}]$, respectively. For the recovery of the support of $\boldsymbol{U}$ from $\boldsymbol{V}$ in the CTF block we apply the SOMP algorithm from [134].

Central frequency and relative autocorrelation estimation We begin with investigating the recovery performance with respect to the SNR. Figure 7.2 shows the frequency support recovery rate (SRR), defined as an average ratio of the intersection size between the true and the estimated frequency supports to $|\mathcal{S}| = K$ (see (5.12)), and the average normalized MSE between the true $\bar{r}_k^{(p_1,p_2)}[\tau_\nu]$ and the estimated $\hat{\bar{r}}_k^{(p_1,p_2)}[\tau_\nu]$ relative autocorrelation functions calculated as $\frac{1}{K(N_{\mathrm{T}}+1)} \sum_{k=1}^{K} \sum_{\nu=0}^{N_{\mathrm{T}}} \frac{|\bar{r}_k^{(p_1,p_2)}[\tau_\nu] - \hat{\bar{r}}_k^{(p_1,p_2)}[\tau_\nu]|^2}{|\bar{r}_k^{(p_1,p_2)}[\tau_\nu]|^2}$ for $N_{\mathrm{T}} = 500$ samples (sensing time of 18 μc). We can see that the proposed recovery methods provide comparable

performance in terms of both frequency and relative autocorrelation estimation. However, the SRR is somewhat higher for the two-step approach, while the joint estimation method provides slightly better accuracy in terms of $\bar{r}_k^{(p_1,p_2)}[\tau_\nu]$ recovery.

In the second experiment, we keep the SNR fixed at 0 dB and change the sensing time N_T. The results for the support recovery rate and the MSE are shown in Figure 7.3. Here we have a similar tendency, namely that the joint estimation outperforms the two-step approach with respect to the accuracy of the relative autocorrelation estimation. On the other hand, the latter is more robust with respect to the support recovery and, as expected, less sensitive to the choice of the sensing time.

7.3 In-band interference detection

Note that in order to be able to use $\bar{r}_k^{(p_1,p_2)}[\nu T_\mathrm{s}]$ for the localization of the kth signal source, we have to ensure that the signals $\bar{x}_{k,p_1}[t_n]$, $\bar{x}_{k,p_2}[t_n]$ observed by sensors p_1, p_2 in the (ℓ_k)th sub-band describe the same emitter. If this is not the case then the composition of $\bar{r}_k^{(p_1,p_2)}[\nu T_\mathrm{s}]$ does not convey any information about the respective emitter locations and, in the best case, can not contribute to the localization process. In the worst case scenario, it might yield ghost paths and erroneous estimates. This said, being able to detect such an in-band interference beforehand would allow the localization system to separate the sensor data into separate sets according to the transmitters observed by different sensors and, subsequently, improve the localization performance. In the following we show that the sub-Nyquist sampling framework considered here provides an opportunity to do so, provided that the number of interfering sources is relatively low.

7.3.1 Scenario description

Suppose that at any moment $N_\mathrm{oc} \leq K$ of N communication sub-bands are occupied by K independent uncorrelated narrowband transmissions $x_k(t)$ and consider a sensing scenario illustrated in Figure 7.4. Due to the propagation conditions each sensor observes $K_\mathrm{obs}^{(p)} \leq K$ transmissions that occupy $K_\mathrm{obs}^{(p)}$ distinct communication sub-bands, meaning that $K - K_\mathrm{obs}^{(p)} \geq 0$ transmissions are hidden from the pth sensing node, e.g., behind an obstacle. To this end, we define $\mathcal{K}_p^\mathrm{o}$ and $\mathcal{K}_p^\mathrm{h}$ as the sets of indices corresponding to the source signals $x_p(t)$ observed and hidden from the pth sensor, respectively, i.e.,

$$\mathcal{K}_p^\mathrm{o} = \{k \in [1, K] : H_{k,p}(f) \neq 0\}, \tag{7.27}$$

$$\mathcal{K}_p^\mathrm{h} = \{k \in [1, K] : H_{p,k}(f) \equiv 0\}, \tag{7.28}$$

where $H_{k,p}(f)$ denotes the frequency response of the propagation channel between the source of kth narrowband transmission and the pth sensor. Additionally, we denote by $\mathcal{C}_p^\mathrm{o}$ the set of indices corresponding to the central frequencies of the occupied communication sub-bands, as

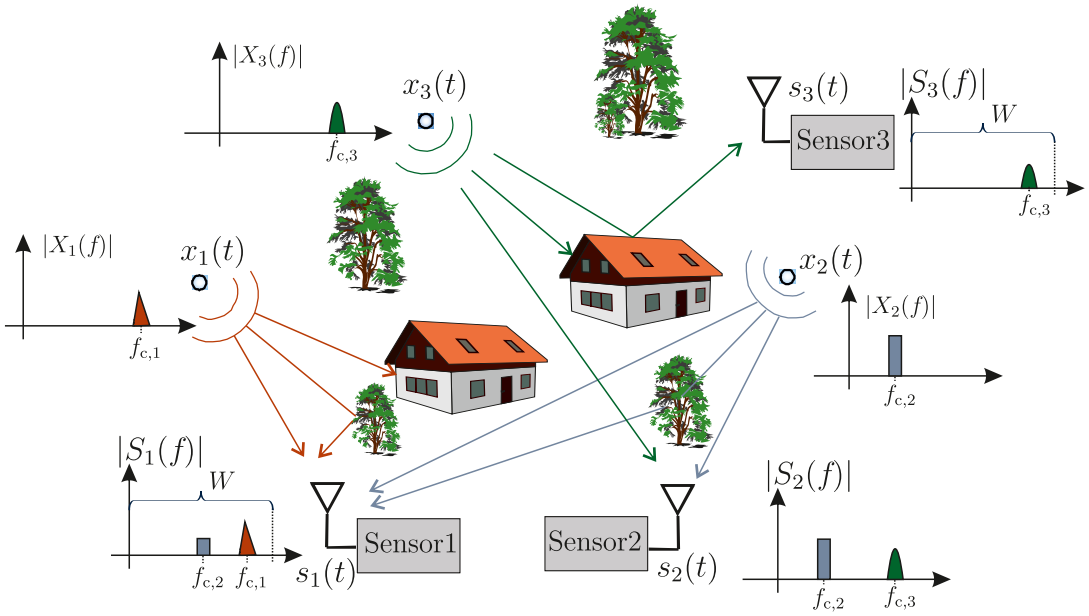


Figure 7.4: An example of a sensing scenario with $P = 3$ sensing nodes observing $K = 3$ transmissions. The transmissions $x_p(t)$ occupy $K_{\text{oc}} = 2$ communication sub-bands such that signals $x_1(t)$ and $x_3(t)$ are potentially interfering with each other.

observed at the pth sensing node such that $\mathcal{C}_p^{\text{o}} = \{n \in [1, N] : f_{\text{c},k} = f_n^{\text{c}} \forall k \in \mathcal{K}_p^{\text{o}}\}$. Finally, we assume that each transmission is received by at least one sensor such that $\bigcup_{p=1}^{P} \mathcal{K}_p^{\text{o}} = \{k\}_{k=1}^{K}$ and $|\bigcup_{p=1}^{P} \mathcal{C}_p^{\text{o}}| = N_{\text{oc}}$ and introduce a following definition.

Definition 12. *We define set $\mathcal{I}$ as the set of all central frequencies $f_n^{\text{c}}, n \in [1, N]$ that correspond to the communication sub-bands occupied by more than one source signal $x_k(t)$, where $k \in [1, K]$.*

According to the above definition, we consider the observed band $\mathcal{F}$ interference free if the set $\mathcal{I}$ is empty. Contrarily, we say that an ith channel, i.e., a channel with a central frequency f_i^{c}, is experiencing in-channel interference if $f_i^{\text{c}} \in \mathcal{I}$. This being said, our goal is to detect the presence of potential interference between the transmitters and to estimate the central frequencies of the communication channels in which it occurs from the sub-Nyquist samples of $s_p(t)$.

7.3.2 Cooperative interference detection from sub-Nyquist samples

7.3.2.1 Covariance matrix analysis

Intra-sensor covariance. Consider the intra-sensor covariance matrix $\boldsymbol{Q}_n$ defined as

$$\boldsymbol{Q}_p = \sum_{n=-\infty}^{\infty} \boldsymbol{y}_p[t_n]\boldsymbol{y}_p^{\text{H}}[t_n] = \int_{f \in \mathcal{F}_s} \boldsymbol{y}_p(f)\boldsymbol{y}_p^{\text{H}}(f)\text{d}f = \boldsymbol{\Phi}\boldsymbol{\Sigma}_p\boldsymbol{\Phi}^{\text{H}}, \tag{7.29}$$

where $\boldsymbol{\Sigma}_p = \int\limits_{f\in\mathcal{F}_s} \boldsymbol{z}_p(f)\boldsymbol{z}_p^{\mathrm{H}}(f)\mathsf{d}f$. Since the individual signals $x_k(t)$ are wide-sense stationary and independent from each other, we have that

$$\boldsymbol{\Sigma}_p = \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{(p)} = \sum_{k=1}^{K} \int\limits_{f\in\mathcal{F}_s} \boldsymbol{z}_k^{(p)}(f)\left(\boldsymbol{z}_k^{(p)}\right)^{\mathrm{H}}(f)\mathsf{d}f, \tag{7.30}$$

where $\boldsymbol{\Sigma}_p^{(n)} = \int\limits_{f\in\mathcal{F}_s} \boldsymbol{z}_k^{(p)}(f)\left(\boldsymbol{z}_k^{(p)}\right)^{\mathrm{H}}(f)\mathsf{d}f$ with $\boldsymbol{z}_k^{(p)}(f)$ being a vector of length L with ℓth element given by

$$\boldsymbol{z}_{k,\ell}^{(p)}(f) = X_{k,p}(f-\ell f_{\mathrm{p}}) = H_{k,p}(f-\ell f_{\mathrm{p}})X_k(f-\ell f_{\mathrm{p}}), f\in\mathcal{F}_s. \tag{7.31}$$

Here, $X_{k,p}(f)$ and $X_k(f)$ stand for the Fourier transforms of $x_{k,p}(t)$ and $x_k(t)$, respectively.

From (7.27), $\boldsymbol{z}_k^{(p)}(f\in\mathcal{F}_{\mathrm{s}})$ has a single non-zero for any $k\in\mathcal{K}_p^{\mathrm{o}}$ and hence we have that $\forall k\in\mathcal{K}_p^{\mathrm{o}}$ the covariance matrix $\boldsymbol{\Sigma}_k^{(p)}$ is a diagonal matrix with a single non-zero element on its main diagonal. Contrarily, for any $k\in\mathcal{K}_p^{\mathrm{h}}$ vector $\boldsymbol{z}_k^{(p)}$ contains only zeros resulting in an all-zero matrix $\boldsymbol{\Sigma}_k^{(p)}$. Therefore, we can rewrite (7.30) as

$$\boldsymbol{\Sigma}_p = \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{(p)} = \sum_{k\in\mathcal{K}_p^{\mathrm{o}}} \boldsymbol{\Sigma}_k^{(p)}, \tag{7.32}$$

which means that $\boldsymbol{\Sigma}_p$ is a diagonal matrix with $\kappa_p = |\mathcal{K}_p^{\mathrm{o}}| = K_{\mathrm{obs}}^{(p)}$ non-zeros on its main diagonal. Obviously, the support of $\boldsymbol{\Sigma}_p$, i.e., the positions of the non-zero elements on its main diagonal, coincides with the support $\mathcal{S}_p$ of $\boldsymbol{z}_p(f\in\mathcal{F}_{\mathrm{s}})$, which can be found by applying the CTF block to (7.29) for instance.

Inter-sensor covariance Let $\boldsymbol{Q}_{p_1,p_2}$ denote an inter-sensor covariance matrix between the outputs of (p_1)th and (p_2)th sensors, defined as

$$\boldsymbol{Q}_{p_1,p_2} = \sum_{n=-\infty}^{\infty} \boldsymbol{y}_{p_1}[t_n]\boldsymbol{y}_{p_2}^{\mathrm{H}}[t_n] = \int\limits_{f_\in\mathcal{F}_s} \boldsymbol{y}_{p_1}(f)\boldsymbol{y}_{p_2}^{\mathrm{H}}(f)\mathsf{d}f = \boldsymbol{\Phi}\int\limits_{f\in\mathcal{F}_s} \boldsymbol{z}_k^{(p_1)}(f)\left(\boldsymbol{z}_k^{(p_2)}\right)^{\mathrm{H}}(f)\mathsf{d}f\boldsymbol{\Phi}^{\mathrm{H}}. \tag{7.33}$$

Similarly to (7.29), we can write $\boldsymbol{Q}_{p_1,p_2}$ as

$$\boldsymbol{Q}_{p_1,p_2} = \boldsymbol{\Phi}\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{(p_1,p_2)}\boldsymbol{\Phi}^{\mathrm{H}} = \boldsymbol{\Phi}\boldsymbol{\Sigma}_{p_1,p_2}\boldsymbol{\Phi}_k^{\mathrm{H}}, \tag{7.34}$$

where $\boldsymbol{\Sigma}_{p_1,p_2} = \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{(p_1,p_2)}$ while $\boldsymbol{\Sigma}_k^{(p_1,p_2)} = \int\limits_{f\in\mathcal{F}_s} \boldsymbol{z}_k^{(p_1)}(f)\left(\boldsymbol{z}_k^{(p_2)}\right)^{\mathrm{H}}(f)\mathsf{d}f$.

Taking into account that $\boldsymbol{z}_k^{(p_1)}(f) = \mathbf{0}\ \forall k \in \mathcal{K}_{p_1}^{\mathrm{h}}$ whereas $\boldsymbol{z}_k^{(p_2)}(f) = \mathbf{0}\ \forall k \in \mathcal{K}_{p_2}^{\mathrm{h}}$, we have that $\boldsymbol{\Sigma}_k^{(p_1,p_2)}$ is non-identically zero if and only if $k \in \mathcal{K}_{p_1,p_2}^{\mathrm{o}} = \mathcal{K}_{p_1}^{\mathrm{o}} \cap \mathcal{K}_{p_2}^{\mathrm{o}}$. Therefore, we immediately obtain that

$$\boldsymbol{\Sigma}_{p_1,p_2} = \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{(p_1,p_2)} = \sum_{k \in \mathcal{K}_{p_1,p_2}^{\mathrm{o}}} \boldsymbol{\Sigma}_k^{(p_1,p_2)} \tag{7.35}$$

is a diagonal matrix with $\kappa_{p_1,p_2} = |\mathcal{K}_{p_1,p_2}^{\mathrm{o}}| \leq \min(\kappa_{p_1}, \kappa_{p_2})$ non-zero elements on its main diagonal. Clearly, the support $\mathcal{S}_{p_1,p_2}$ of $\boldsymbol{\Sigma}_{p_1,p_2}$ is a subset of the intersection between the individual supports $\mathcal{S}_{p_1}$ and $\mathcal{S}_{p_2}$, i.e., $\mathcal{S}_{p_1,p_2} \subseteq (\mathcal{S}_{p_1} \cup \mathcal{S}_{p_2})$.

Example 7.21

In this example we illustrate the composition of the intra- and inter-sensor signal covariance matrices $\boldsymbol{\Sigma}_p$, $\boldsymbol{\Sigma}_{p_1,p_2}$ for the sensing scenario from Figure 7.4 where $p = 1, 2, 3$ while $p_1 = 1, 3$ and $p_2 = 2, 3$. In the upper row of Figure 7.5, we present schematic examples of $\boldsymbol{\Sigma}_p$, while $\boldsymbol{\Sigma}_{p_1,p_2}$ are shown in its lower row. The individual transmissions in Figure 7.5 are represented by different colors: $x_1(t)$ by red, $x_2(t)$ by green, and $x_3(t)$ by blue. We observe the difference between the supports of inter- and cross-covariance matrices: the inter-covariance matrices retain non-zero entries only at the indices corresponding to the locations of sub-bands that have no in-band interference.

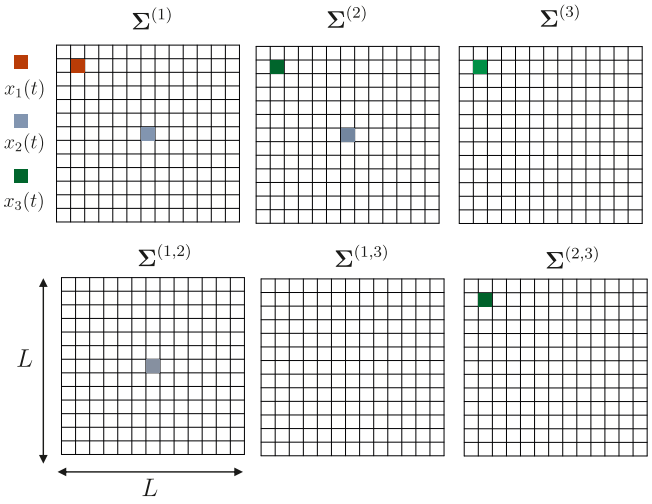


Figure 7.5: An example of intra-sensor (upper row) and inter-sensor (lower row) signal covariance matrices for the sensing scenario from Figure 7.4.

7.3.2.2 Centralized interference detection

Having determined the structure of the covariance matrices, we now turn to showing how one can exploit it for the purpose of in-band interference detection. Denote by $\boldsymbol{\sigma}_p = \mathrm{diag}\{\boldsymbol{\Sigma}_p\}$the main diagonal of the pth inter-sensor covariance matrix and consider the following lemma.

Lemma 4. *Let $\mathcal{S}_{p_1}$ and $\mathcal{S}_{p_2}$ be the supports of $\boldsymbol{\sigma}_{p_1}$ and $\boldsymbol{\sigma}_{p_2}$, respectively. Then, if $\mathcal{S}_{p_1} \neq \mathcal{S}_{p_2}$ the two sets $\mathcal{K}_{p_1}^{(\mathrm{o})}$ and $\mathcal{K}_{p_2}^{(\mathrm{o})}$ are non-identical.*

Proof. Cf. Appendix B.5.1. ■

Lemma 4 states that if for some pair of sensing nodes the supports recovered at each node are not identical, then the sets of transmissions observed by these nodes are not identical either. However, the knowledge of the individual supports is not enough to detect potential in-band interference. To illustrate this, consider examples of intra-sensor signal covariance matrices in Figure 7.5 of Example 7.21. As can be seen from $\mathbf{\Sigma}^{(1)}$ and $\mathbf{\Sigma}^{(2)}$ the supports $\mathcal{S}_1$ and $\mathcal{S}_2$ are equal, but the sets of source signals observed by the two sensors are different. To detect this situation, we propose to additionally estimate the inter-sensor supports.

Denote by $\boldsymbol{q}_{p_1,p_2} = \mathrm{vec}\,(\boldsymbol{Q}_{p_1,p_2})$ a length-M^2 vector that contains the columns of $\boldsymbol{Q}_{p_1,p_2}$ stacked one under another. Then, since $\mathbf{\Sigma}_{p_1,p_2}$ is a diagonal matrix, we can write that

$$\boldsymbol{q}_{p_1,p_2} = (\mathbf{\Phi}^* \otimes \mathbf{\Phi})\, \boldsymbol{r}_{p_1,p_2} = \big(\mathbf{\Phi}^* \odot \mathbf{\Phi}\big)\boldsymbol{\sigma}_{p_1,p_2} = \mathbf{\Psi}\boldsymbol{\sigma}_{n,k}, \tag{7.36}$$

where $\boldsymbol{r}_{p_1,p_2} = \mathrm{vec}\,(\mathbf{\Sigma}_{p_1,p_2})$, $\boldsymbol{\sigma}_{p_1,kp_2} = \mathrm{diag}\,(\mathbf{\Sigma}_{p_1,p_2})$ is a κ_{p_1,kp_2}-sparse vector of length L and $\mathbf{\Psi} = \big(\mathbf{\Phi}^* \diamond \mathbf{\Phi}\big)$. Solving (7.36) for the sparsest $\boldsymbol{\sigma}_{p_1,p_2}$, we find the support[5] $\mathcal{S}_{p_1,p_2}$. Given (7.36), we now provide the conditions for the detection of in-channel interference from $\mathcal{S}_{p_1,p_2}$.

Theorem 9. *Denote by $\times$ the Cartesian set product and let $\mathcal{S}_p$, $\mathcal{S}_{p_1,p_2}$ be the supports of $\boldsymbol{\sigma}_{p_1}$, $\boldsymbol{\sigma}_{p_1,p_2}$, respectively. Then, the set $\mathcal{I}$ is non-empty if and only if there exists a pair $(p_1, p_2) \in \mathcal{N}_\mathrm{q} = \bigcup_{q=2}^{P} \{[1, q-1] \times q\}$ such that $\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2} \neq \emptyset$ and $\mathcal{S}_{p_1,p_2} \neq \mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}$.*

Proof. Cf. Appendix B.5.2. ■

Theorem 9 states that the presence of in-band interference can be detected by estimating the supports of $\mathbf{\Sigma}_{p_1,p_2}$, $\mathbf{\Sigma}_p$ and comparing them with each other. Once the interference is detected, we can also estimate the central frequencies of the communication sub-bands in which it occurs by completing the set $\mathcal{I}$ as described in the following corollary.

Corollary 2. *Let $\bar{\mathcal{I}}_{p_1,p_2} = (\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}) \setminus \mathcal{S}_{p_1,p_2}$ be the relative compliment of $\mathcal{S}_{p_1,p_2}$ in $\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}$. Given the set of supports $\{\mathcal{S}_p\}_{p=1}^{P}$ and $\{\mathcal{S}_{p_1,p_2}\}_{(p_1,p_2)\in\mathcal{N}_\mathrm{q}}$, the set of central frequencies that are occupied by more than one source signals can be found as*

$$\mathcal{I} = \{\ell_i f_\mathrm{p} : \ell_i \in \bar{\mathcal{I}}\}, \tag{7.37}$$

where $\bar{\mathcal{I}} = \bigcup_{(p_1,p_2)\in\mathcal{N}_\mathrm{q}} \bar{\mathcal{I}}_{p_1,p_2}$.

7.3.3 Performance evaluation

We now validate the proposed in-band interference detection approach by performing numerical simulations.

[5]Note that when all sensing nodes use the same mixing sequences (the sampling matrices $\mathbf{\Phi}$ are the same for all $p \in [1, P]$) the support $\mathcal{S}_{p_1,p_2}$ can be also recovered by using the CTF block. However, when this is not the case and the sensing matrices (i.e., the sensing sequences) differ solving (7.36) is the only option.

7.3.3.1 Simulations setup

Signal and sensing parameters. We consider $K = 3$ BPSK modulated signals $x_k(t)$ with a bandwidth $B = 20$ MHz each that are distributed over a frequency band of $W = 4$ GHz. The band W is split into $N = 194$ communication channels such that the central frequencies $f_{\mathrm{c},k}$ are chosen from the set $\{nW/N\}_{n=1}^{N}$. The number of transmitters K_{i} that occupy the same channel is generated according to the Bernoulli distribution with $\rho = 0.5$. On the sensing side, we consider $P = 3$ sensing nodes that operate with $M = 20$ sampling channels at the sampling rate of $f_{\mathrm{s}} = 21$ MHz each. The mixing sequences $p_m(t)$ in (7.9) are generated as pseudo-random $\{\pm 1\}$ piece-wise constant functions. The total sampling rate at each sensor is 420 MHz which is $\sim 5\%$ of the Nyquist rate, whereas the total sensing time is 4.5 μs or equivalently 100 samples.

Propagation parameters. Each signal passes through a propagation channel where it experiences path-loss and log-normal shadowing. The channel impulse response in this case is a constant given by

$$h_{k,p} = \beta_{k,p} 10^{-PL_{k,p}/20}, \tag{7.38}$$

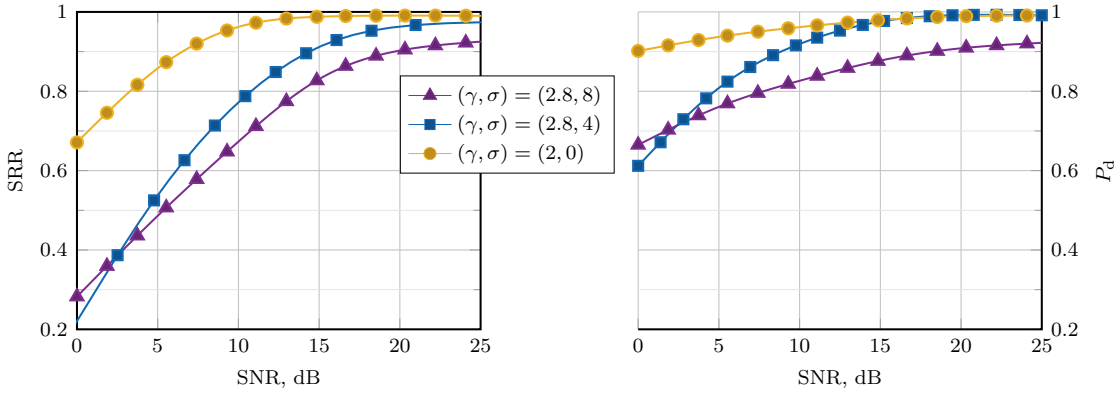
where $PL_{k,p}$ is the total path loss between the kth signal source and the pth sensor. The coefficient $\beta_{k,p}$ in (7.38) is equal to zero if the kth signal is completely blocked from the pth sensor and is set to 1 otherwise. Measured in dB, the average path loss is expressed as

$$PL_{k,p} = PL_0 + 10\gamma \log_{10}(\frac{d_{k,p}}{d_0}) + X_\sigma, \tag{7.39}$$

where the d_0 is the distance to the reference point in the far-field of the receiver antenna and PL_0 is the path loss at the point corresponding to d_0. Without loss of generality, we set these to 0 and 1, respectively. Distances between the kth signal source and the pth receiver $d_{k,p}$ are generated uniformly at random between 1 and 5 km. The parameter γ in (7.39) is the path loss exponent and X_σ is a zero-mean Gaussian random variable with variance σ^2 that reflects the attenuation caused by the large-scale shadowing. In case of no shadowing, i.e., $X_\sigma \equiv 0$, PL becomes a linear function of the distance $d_{k,p}$. The values of γ and σ are usually determined empirically; their typical ranges in outdoor environments are estimated to be $[2, 4]$ and $[4, 10]$ dB, respectively [325], [326]. Finally, the sets $\mathcal{K}_p^{\mathrm{o}}$ and $\mathcal{C}_p^{\mathrm{o}}$ are constructed such that κ_p is generated uniformly at random from $[1, 3]$ if $N_i = 0$ and from $[1, 2]$ if $N_i = 1$, while $\forall k \in [1, K]$ there exists $\beta_{k,p} \neq 0$.

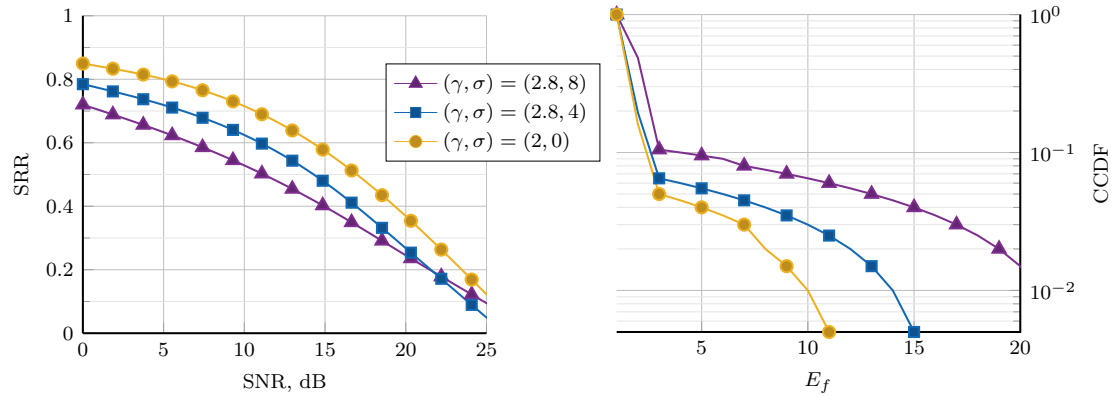
7.3.3.2 Performance evaluation

We begin by examining the detection performance with respect to the SNR. Figure 7.6 presents average intra-sensor support recovery rate (SRR) (Figure 7.6a) and the probability of correct interference detection P_{d} (Figure 7.6b), calculated as the ratio between the number of iterations when the set $\mathcal{I}$ was correctly detected non-empty to the total number of realizations for several combinations of propagation parameters[6] (γ, σ). The probability of detection grows proportionally with the average support recovery rate, since it depends on how accurately the intersection between a pair of individual supports is estimated. It can be observed however that P_{d} is on average greater than SRR. This reflects the fact that it is not necessary to estimate all the individual supports correctly in order to detect the presence of the interference. For comparison, Figure 7.7a demonstrates the corresponding false alarm rate P_{fa} computed as a percentage of cases when a channel has been detected as containing interfering sources while in reality it did not.



(a) SRR **(b)** P_{d}

Figure 7.6: Probability of correct interference detection P_{d} (a) and intra-sensor support recovery rate (b).



(a) P_{fa} **(a)** CCDF of E_f

Figure 7.7: Probability of false interference detection P_{fa} (a) and CCDF of the normalized frequency estimation error E_f for an SNR of 20 dB (b).

[6]Note that the parameter combination $(2, 0)$ corresponds to the free-space propagation with no shadowing.

Finally, in Figure 7.7b we evaluate the central frequency estimation performance by computing the complimentary cumulative distribution function (CCDF) of the normalized frequency estimation error

$$E_f = \frac{1}{N_{\text{it}}} \sum_{N_{\text{it}}} \frac{|(\hat{\ell}_i - \ell_i)W/N|}{W/N} = \frac{1}{N_{\text{it}}} \sum_{N_{\text{it}}} |\hat{\ell}_i - \ell_i|,$$

where $N_{\text{it}} = 10^3$ is the number of (noise) iterations and $\hat{\ell}_i \in \hat{\mathcal{I}}$ while $\ell_i \in \mathcal{I}$. We see that the majority of the estimates concentrate around zero for all presented cases, while the spread of the erroneous estimates grows with the increase of the average path loss and the decrease of the SNR.

7.4 Concluding remarks

In this chapter, we focused on the tasks of multiband spectral sensing and time delay estimation from sub-Nyquist samples. Assuming a network of distributed time-synchronized sub-Nyquist sensing nodes, we developed a method to estimate relative autocorrelation functions of the individual signals within the active sub-bands from low rate outputs of different sensors. More specifically, we proposed two alternative estimation methods that differ by the way the relative autocorrelation functions are computed. Our numerical results demonstrate that both proposed methods allow estimating spectral occupancy and the relative autocorrelation functions of the individual signals from the sub-Nyquist samples of the received wideband signals. As expected, the joint estimation method tends to be somewhat superior in terms of the accuracy of the relative autocorrelation estimation, while the two-step approach provides a higher frequency recovery rate. Once the relative autocorrelation functions of the individual signals are obtained one could determine the corresponding TDoAs and proceed with emitter location estimation using any of the available time delay based localization techniques. This however, is only possible when the estimated relative autocorrelation functions describe the same source locations. As mentioned at the beginning of this chapter, due to varying propagation conditions it is possible that different sensors observe different source signals in the same band, while being unaware of this. Thankfully, the structure of the covariances between the low-rate outputs of different sensors acquired in the considered sub-Nyquist sampling framework provides an opportunity to detect the presence of such an in-band interference. This said, in the second part of this chapter we developed an approach for centralized in-band interference detection in a distributed network of sub-Nyquist sensors.

Chapter 8

Conclusions and outlook

In this thesis, we studied the operation of a compressive sensing system in a generic setting as well as in an application to multiband sensing with the main goal of increasing our understanding of the CS-based signal acquisition process and the implications it has to the practical sub-Nyquist sampling and processing of continuous analog signals. With this in mind, in the first part of this thesis we explored some of the basic elements of the CS framework to assess their interplay and impact on the system performance. We began with investigating the influence of the choice of the measurement kernel on the system parameters and the recovery performance. By looking at the sensing matrix from a deterministic perspective, rather than taking a commonly assumed asymptotic view, we discovered that not only does the application of a compressive measurement kernel potentially affects the characteristics of the input (signal) noise [80], [155] but it also causes variations of the effective signal power in the compressed measurements over the signal support in the sparsifying representation. As a result, for a fixed input SNR, the effective SNR at the output of the receiver becomes dependent on the support of the input signal; a situation that does not occur in traditional Nyquist rate sensing. We analyzed the degree of such variations for several common sensing matrix choices and examined its impact on the recovery performance, which showed that the spread of the output SNR results in the support-dependent recovery performance where the worst- and the best-case performance largely corresponds to the supports that yield the lowest and the highest effective SNRs. An understanding of this phenomenon and its implications for sensing performance is especially important for practical implementations of CS-based acquisition hardware: the measurement kernel in this case is likely to have equivalent dimensions that are far from the asymptotic regime, while its elements might need to be fixed (at least temporarily) in order to be realized in hardware. This also highlights the need for alternative measurement design strategies that would provide more control over system parameters, such as the degree of output SNR variations, and could accommodate different application-specific requirements. Inspired by the promising results of the sensing matrix optimization based on the minimization of the matrix coherence [156], [327]–[329], in this thesis we introduced a design framework that can be used to optimize

the measurement kernel according to a particular set of system requirements, such as the minimization of the CRB or a certain structure of the column correlations, as in direction finding (DF) applications for instance [A3], [A6]. While the analysis presented in this thesis sheds light on several crucial aspects of the compressive measurement kernel design, it also reveals the importance of further investigations in this area. For compressive sensing systems to see a practical implementation, it is paramount that there is a solid understanding of the relation between the measurement design and the performance it provides. This includes not only the design itself but also its sensitivity to different types of perturbations [A5], caused by hardware limitations and/or implementation constraints, development of calibration procedures, etc.

From analyzing the sensing matrix we turned to examining another founding premise of the CS theory, namely that of the1 sparse structure of the input signal. Surely, any (bandlimited) signal can be sub-sampled, e.g. by simply reducing the sampling rate below what is required by its natural representation. However, not every signal can then be recovered from such reduced-rate measurements. For generic signals, simply reducing the sampling rate would likely result in an irreversible partial loss of information, which, depending on the degree of sub-sampling, could lead to completely meaningless measurements. What distinguishes the types of signals that CS systems target is that despite being naturally represented in a seemingly large-dimensional space, they belong to some lower-, albeit unknown, dimensional sub-spaces. In other words, the signals we wish to acquire are sparse (or compressible): they can be described by only a few non-zero coefficients and a corresponding number of generating functions. The degree of how sparse the signal is, which we refer to as the sparsity order, has a tremendous impact on system design and performance. It determines how many measurements are required for successful recovery and indicates what recovery performance one can expect. In this thesis, we proposed an approach for sparsity order estimation (SOE) where we infer the sparsity order from the rank of a properly arranged matrix of (compressive) measurements. Building upon the ideas from principal component analysis and model order selection (MOS), it allows us to infer the order directly from the compressed domain, without the need to perform multiple signal reconstructions or sensing matrix modifications. We showed that the proposed SOE approach can be applied to both of the most common CS signal models, namely the multiple measurement and the canonical single measurement vector models, where for the latter we ensure the possibility of order estimation by introducing a specific design of the sensing matrix. Besides being beneficial to some of the reconstruction algorithms, the knowledge of the sparsity order could allow us to adapt our reconstruction strategy, e.g., by choosing a recovery algorithm whose performance and complexity is best suited to the current signal's sparsity. It also offers an opportunity to adapt the measurement to the complexity of the current signal (scene), provided that the application allows for it. These and other potentials of the sparsity order estimation are awaiting further exploration.

Having explored the relations between the components of a generic CS system, we moved to investigating the application of the CS paradigm to sub-Nyquist sampling and processing of analog continuous signals. In this thesis, we focused on a particular example, namely that of sensing analog wideband signals with a multiband structure. The input signal with a multiband spectrum is characterized by the presence of multiple narrowband communication sub-bands within the total signal's bandwidth. These sub-bands are usually assumed to be occupied by independent transmissions whose number is relatively low compared to the total number of the available sub-bands. Given such a wideband multiband signal, we acquire it at a sub-Nyquist rate by using the modulated wideband converter (MWC) receiver architecture [254], [255], which practically implements sub-Nyquist sampling in the union of subspaces [146], [254]. The MWC is a multi-channel receiver system where in each channel the wideband input first undergoes the process of analog bandwidth reduction, after which it is sampled at a low sampling rate. The overall sampling rate in the MWC is thus defined by the total number of sampling channels times the sampling rate per channel.

With this in mind, the first task we considered is the multiband spectrum sensing from sub-Nyquist samples. Having been a topic of active research for a number of years by now, the approaches to multiband spectrum sensing from sub-Nyquist samples seem to have largely focused on the recovery of the full signal spectrum or the PSD, which would then be passed to one of the available (Nyquist-rate) multiband detection schemes. Typically, these provide a coarse estimation of the spectral occupancy at the resolution of the adopted multiband detector. Contrary, in this thesis we proposed a compressive multiband detector that avoids performing the step of spectral/PSD reconstruction by directly recovering the sub-band energies at the coarse resolution of the spectral channelization of the MWC. Besides requiring fewer measurements due to the discarding of the phase information, it also provides an opportunity to partially solve the problem of selecting an adequate threshold common to traditional energy detection methods. One of the main disadvantages of the proposed detector, however, is its fixed spectral resolution that is defined by the hardware implementation rather than by tuning a digital filterbank. Nevertheless, if the coarse occupancy estimation is not enough, one can still recover the PSD from the same samples and perform sensing refinement using the coarse estimation as an initial guess. Finally, in light of the sensing matrix analysis performed in the generic CS setting, we briefly examined the composition of the sensing matrix in the MWC and its connection to the detection performance, which demonstrated a familiar trend: the application of the compressive sampling kernel results in the variations of the effective output SNR that translates into the support-dependent detection performance. The influence of this effect in the MWC, however, calls for further investigation, as the sensing matrix can have different structures depending on the choice of the parameters at the bandwidth reduction step, while its interaction with the input signal has a somewhat more intricate character.

After considering sub-Nyquist sensing in frequency, we extended the sensing task to include the angular domain by equipping the sensor with an antenna array and developing a

sub-Nyquist receiver system capable of angular-frequency spectrum recovery. Keeping the MWC as the underlying sub-Nyquist sampling architecture, we sample each antenna output with $M \geq 1$ MWC channels which yields the total number of receiver channels equal to MN, where N stands for the number of antenna elements. Having formulated the task of angular-spectral sensing from the obtained sub-Nyquist samples in the CS framework, we derived conditions on the total number of sampling channels versus the number of antennas for perfect recovery of the 2D angular-frequency spectrum and provided approaches for estimating the central frequencies and the DoAs of the signals within the active sub-bands. Our analysis indicates that for non-overlapping in frequency sources, the number of antenna elements can be traded (to some extent) for the number of MWC channels per antenna. In a noisy environment, this trade-off is however asymmetrical in the sense that it tends to favor antennas over the MWC channels. This can be explained by noting that the system noise is likely to be dominated by the noise at the output of the antennas (e.g., generated by the low noise amplifiers (LNAs) or as a part of the interference level) that is processed by the sampling system together with the wideband input. In this case, the noise at the different low-rate outputs of the same antenna will be correlated, while the noise at the outputs of different antennas is likely to have independent statistics. Contrary to the related systems proposed in the literature, the angular-spectral sensing approach described in this thesis does not rely on a specific structure of the antenna array responses, such as uniform sensor spacing or isotropic elements. In contrast, it can be applied to (almost) arbitrary antenna arrays and allows the inclusion of full 3D polarimetric array responses to reflect the sensitivity of the antennas with respect to both azimuth and elevation angles of arrival, as well as the polarization states of the incoming waves, which is particularly important for applications with realistic antenna arrays. This being said, the presented analysis initiates next steps towards a more practical sub-Nyquist multiband antenna system for angular-spectral sensing, while a large number of important research questions remains to be answered. The most crucial questions are concerned with the design and realistic modelling of wideband antenna arrays capable of multiband operation in the sub-Nyquist setting, as well as the interplay between the antenna array and the MWC characteristics in a practical implementation. Another block of questions is related to the development of estimation approaches that could more effectively tackle multi-dimensional signal structure without resorting to building large two-dimensional dictionaries. Finally, there is the issue of system calibration, that is crucial both for the array and the MWC operation.

In all of the above, we considered sub-Nyquist sensing at a single isolated sensor. Obviously, the discussed sensing tasks can also be performed in a network of spatially distributed sensing nodes that cooperate by exchanging the measurement data with each other or a centralized processing unit. By taking advantage of spatial diversity, distributed cooperative sensing offers the potential of the improved detection performance and increased robustness to propagation conditions [330]–[332]. On the other hand, having a distributed network of sensors capable of DF provides an opportunity for source localization, provided that the sensors know their

relative positions [333]. In this thesis, we focused on a complementary approach to source localization based on relative time delay estimation. Assuming a network of distributed time-synchronized sub-Nyquist sensing nodes, we developed a method to estimate relative autocorrelation functions of the individual signals within the active sub-bands of a multiband input from its sub-Nyquist samples obtained at different sensors. More specifically, we proposed two alternative estimation methods, both of which are based on spectral support recovery. However, to be able to use the estimated relative autocorrelation functions for source localization one has to ensure that they describe the same source locations. Note that in the distributed scenario, it is possible that different sensors can observe different source signals in the same band. This can happen due to the peculiarity of the propagation conditions or even due to malicious intent. With this said, being able to detect such in-band interference beforehand would allow the localization system to separate the sensor data according to the set of transmitters observed by different sensors and, subsequently, improve the localization performance. By exploiting the structure of the covariances between the low-rate outputs of different sensors, we demonstrated that in the sub-Nyquist sampling framework this can be done under mild conditions on the number of sensors and the number of interfering sources. Once the relative autocorrelation functions are found, one could determine the corresponding TDoAs and proceed with emitter localization using available approaches, such as trilateration for instance. One could also potentially use the (partial) knowledge of the channel impulse response (obtained from channel measurements, geometry-based modelling and/or ray-tracing tools) to estimate the emitters' locations, using a multilateration method as in [323] for instance. Although a detailed discussion of precise localization techniques is outside the scope of this thesis, one could envisage an extension of the proposed approach to a system that formulates both the spectral detection and the (multiband) localization task as a joint hybrid recovery problem.

In conclusion, the study presented in this thesis provides a number of novel insights into the tasks of compressive acquisition and processing of sparse analog signals, particularly in application to multidimensional multiband sensing. It is our hope that this thesis will encourage further discussion and stimulate the development of the proposed sensing methods towards an implementation in practical sensing systems.

Appendices

Appendix A

Earth mover's distance

A.1 EMD Defintion

Consider two arbitrary vectors $\boldsymbol{p} \in \mathbb{R}^N$ and $\boldsymbol{b} \in \mathbb{R}^N$ and let $\boldsymbol{F}$ be an $N \times N$ non-negative matrix with (i,j) element $f_{i,j} \geq 0$. The EMD between $\boldsymbol{p}$ and $\boldsymbol{b}$, denoted by $d_{\text{EMD}}(\boldsymbol{p}, \boldsymbol{b})$, is then defined as [181]

$$d_{\text{EMD}}(\boldsymbol{p}, \boldsymbol{b}) \triangleq \frac{\min_{\boldsymbol{F}} R(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{F})}{\|\operatorname{argmin}_{\boldsymbol{F}} R(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{F})\|_{1,1}} \text{ s.t. } \begin{cases} \sum_j f_{i,j} \leq |p_i|, \ i = 1, 2, \ldots, N \\ \sum_i f_{i,j} \leq |b_j|, \ j = 1, 2, \ldots, N \end{cases} \tag{A.1}$$

where the objective function $R(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{F})$ is given by

$$R(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{F}) = \sum_{j}^{N} \sum_{i}^{N} f_{i,j} |i - j|. \tag{A.2}$$

In the EMD literature, $R(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{F})$ is usually given the meaning of work required to move mass from $\boldsymbol{p}$ to obtain $\boldsymbol{b}$ [180]. The elements $f_{i,j}$ can be interpreted as the amount of mass that flows from the ith element of $\boldsymbol{p}$ to the jth element of $\boldsymbol{b}$.

A.2 EMD between signal supports

According to (A.1), the EMD is defined on the space of real vectors. In order to apply EMD to the support comparison problem, we introduce a following definition.

Definition 13. *Let $\boldsymbol{v}$ and $\boldsymbol{q}$ be two arbitrary complex vectors of length N. Then, the EMD between two supports $\mathcal{S}(\boldsymbol{v})$ and $\mathcal{S}(\boldsymbol{q})$ is the EMD between two vectors $\tilde{\boldsymbol{v}}$ and $\tilde{\boldsymbol{q}}$ obtained from $\boldsymbol{v}$ and $\boldsymbol{q}$ by setting their non-zero entries to one and normalizing the total vector masses:*

$$d_{\text{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = d_{\text{EMD}}(\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{q}}), \ \textit{where} \tag{A.3}$$

$$\tilde{v}_i = \begin{cases} \frac{1}{K_{\boldsymbol{v}}},\ \textit{if } i \in \mathcal{S}(\boldsymbol{v}) \\ 0,\ \textit{otherwise} \end{cases} \textit{ and } \tilde{q}_j = \begin{cases} \frac{1}{K_{\boldsymbol{q}}},\ \textit{if } j \in \mathcal{S}(\boldsymbol{q}) \\ 0,\ \textit{otherwise} \end{cases}.$$

Due to the mass normalization $\sum_i \sum_j f_{i,j} = 1$ and hence

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \min_{\boldsymbol{F}} R(\boldsymbol{v}, \boldsymbol{q}, \boldsymbol{F}) \text{ s.t. } \begin{cases} \sum_j f_{i,j} = \frac{1}{K_{\boldsymbol{v}}},\ \text{if } i \in \mathcal{S}(\boldsymbol{v}) \\ \sum_i f_{i,j} = \frac{1}{K_{\boldsymbol{q}}},\ \text{if } j \in \mathcal{S}(\boldsymbol{q}) \\ f_{i,j} = 0,\ \text{otherwise} \end{cases}. \tag{A.4}$$

The last constraint in (A.4) is due to the absence of mass flow from or to zero elements. It is useful to re-write (A.4) in matrix form as

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \min_{\boldsymbol{F}} (\mathbf{1}_N^{\mathrm{T}} (\boldsymbol{F} \odot \boldsymbol{G}) \mathbf{1}_N) \text{ s.t. } \begin{cases} \boldsymbol{F}\mathbf{1}_N &= \tilde{\boldsymbol{v}} \\ \boldsymbol{F}^{\mathrm{T}}\mathbf{1}_N &= \tilde{\boldsymbol{q}} \end{cases}, \tag{A.5}$$

where $\boldsymbol{G}$ is an $N \times N$ circulant matrix with an associated polynomial $\sum_{n=0}^{N-1} nx^n$, $\mathbf{1}_N$ denotes an all-one vector of length N and $\odot$ denotes the element-wise matrix product. In the following, we will assume that the elements of $\mathcal{S}(\boldsymbol{x})$ are ordered such that $s_j(\boldsymbol{x}) < s_{j+1}(\boldsymbol{x})\ \forall j \in [1, K_{\mathrm{x}} - 1]$.

A.2.1 Supports with identical cardinalities

Let $K = K_{\boldsymbol{v}} = K_{\boldsymbol{q}}$ and $\boldsymbol{A}$ be a $K \times K$ principal submatrix of $\boldsymbol{F}$ formed by taking the elements indexed by $\mathcal{S}(\boldsymbol{v})$ in column dimension and by $\mathcal{S}(\boldsymbol{q})$ in row dimension, i.e.,

$$\boldsymbol{A} = (\boldsymbol{F}^{\mathrm{T}}_{\mathcal{S}(\boldsymbol{v})})^{\mathrm{T}}_{\mathcal{S}(\boldsymbol{q})}. \tag{A.6}$$

Since all non-zero elements in both $\boldsymbol{v}$ and $\boldsymbol{q}$ have the same mass of $1/K$, we can represent $\boldsymbol{A}$ as $\boldsymbol{A} = \frac{1}{K}\tilde{\boldsymbol{A}}$ where $\tilde{\boldsymbol{A}}$ is comprised of elements $\tilde{a}_{i,j} \in [0, 1]$ with $\sum_i \tilde{a}_{i,j} = \sum_j \tilde{a}_{i,j} = 1$ for all $i, j \in [1, N]$. Then, (A.5) becomes

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \frac{1}{K} \min_{\tilde{\boldsymbol{A}}} (\mathbf{1}_K^{\mathrm{T}} (\tilde{\boldsymbol{A}} \odot \boldsymbol{D}) \mathbf{1}_K) \text{ s.t.} \sum_{i=1}^{N} \tilde{a}_{i,j} = \sum_{j=1}^{N} \tilde{a}_{i,j} = 1, \tag{A.7}$$

where $\boldsymbol{D} = (\boldsymbol{G}^{\mathrm{T}}_{\mathcal{S}(\boldsymbol{v})})^{\mathrm{T}}_{\mathcal{S}(\boldsymbol{q})}$.

Theorem 10. *Let $\boldsymbol{\gamma} \in \mathbb{R}^K$ be a vector that contains some permutation of the $\{1, 2, \ldots, K\}$ sequence and let $\boldsymbol{e}_j$ denote a K-length vector with a single non-zero entry at the jth position.*

Then, the EMD between the supports of two K-sparse vectors $\boldsymbol{v}$ and $\boldsymbol{q}$ can be found as

$$d_{\text{EMD}}(\mathcal{S}(\boldsymbol{v}),\mathcal{S}(\boldsymbol{q})) = \frac{1}{K}\min_{\boldsymbol{\gamma}}(\mathbf{1}_K^{\text{T}}(\boldsymbol{P}_{\boldsymbol{\gamma}} \odot \boldsymbol{D})\mathbf{1}_K), \tag{A.8}$$

where $\boldsymbol{P}_{\boldsymbol{\gamma}} = [\boldsymbol{e}_{\gamma_1}, \boldsymbol{e}_{\gamma_2}, \ldots, \boldsymbol{e}_{\gamma_K}]$ is a $K \times K$ permutation matrix and (n,m)th element of $\boldsymbol{D}$ is given by $d_{n,m} = |s_n(\boldsymbol{v}) - s_m(\boldsymbol{q})|$.

Proof. Consider the minimization problem in (A.7) and note that since $\sum_i \tilde{a}_{i,j} = \sum_j \tilde{a}_{i,j} = 1$, $\tilde{\boldsymbol{A}}$ belongs to the class of doubly stochastic matrices. Then, according to the Birkhoff-von Neumann theorem, we can decompose it as

$$\tilde{\boldsymbol{A}} = \sum_{i=1}^{k} \theta_i \boldsymbol{P}_{\boldsymbol{\eta}_i}, \tag{A.9}$$

where $\boldsymbol{P}_{\boldsymbol{\eta}_i}$ is a $K \times K$ permutation matrix, $k \leq K^2 - 2K + 2$ and $\sum_i \theta_i = 1$ with $\theta_i \in (0,1)$ [334]. Denoting by $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_k]$ and by $\boldsymbol{H} = [\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_k]$, we have that

$$d_{\text{EMD}}(\mathcal{S}(\boldsymbol{v}),\mathcal{S}(\boldsymbol{q})) = \frac{1}{K}\min_{\boldsymbol{\theta},\boldsymbol{H}} \left(\mathbf{1}_K^{\text{T}}(\sum_{i=1}^{k} \theta_i \boldsymbol{P}_{\boldsymbol{\eta}_i} \odot \boldsymbol{D})\mathbf{1}_K \right) \text{ s.t. } \begin{cases} \tilde{\boldsymbol{A}} = \sum_{i=1}^{k} \theta_i \boldsymbol{P}_{\boldsymbol{\eta}_i} \\ \sum_{i=1}^{k} \theta_i = 1 \end{cases}. \tag{A.10}$$

Now, consider the cost function in (A.10):

$$\mathbf{1}_K^{\text{T}} \left(\sum_{i=1}^{k} \theta_i \boldsymbol{P}_{\boldsymbol{\eta}_i} \odot \boldsymbol{D} \right) \mathbf{1}_K = \sum_{i=1}^{k} \theta_i \left(\mathbf{1}_K^{\text{T}}(\boldsymbol{P}_{\boldsymbol{\eta}_i} \odot \boldsymbol{D})\mathbf{1}_K \right) \geq \sum_{i=1}^{k} \theta_i \mu = \mu, \tag{A.11}$$

where $\mu = \min_{\boldsymbol{\gamma}} \left(\mathbf{1}_K^{\text{T}}(\boldsymbol{P}_{\boldsymbol{\gamma}} \odot \boldsymbol{D})\mathbf{1}_K \right)$. Thus, $\tilde{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{\gamma}}$ is an optimum solution to (A.8) and hence

$$d_{\text{EMD}}(\mathcal{S}(\boldsymbol{v}),\mathcal{S}(\boldsymbol{q})) = \frac{1}{K}\min_{\boldsymbol{\gamma}} \left(\mathbf{1}_K^{\text{T}}(\boldsymbol{P}_{\boldsymbol{\gamma}} \odot \boldsymbol{D})\mathbf{1}_K \right). \qquad \blacksquare$$

Theorem 10 states that in order to calculate the EMD between two supports of equal cardinality, it is always enough to move the entire mass from each (non-zero) element of $\boldsymbol{v}$ to only one of each of the (non-zero) elements of $\boldsymbol{q}$. In other words,

$$d_{\text{EMD}}(\mathcal{S}(\boldsymbol{v}),\mathcal{S}(\boldsymbol{q})) = \frac{1}{K}\min_{\boldsymbol{\gamma}} \sum_{n=1}^{K} d_{\gamma_n,n}. \tag{A.12}$$

Since the elements $s_i(\boldsymbol{v})$ and $s_i(\boldsymbol{q})$ are sorted, one can show that

$$d_{\text{EMD}}(\mathcal{S}(\boldsymbol{v}),\mathcal{S}(\boldsymbol{q})) = \frac{1}{K} \sum_{i=1}^{K} |s_i(\boldsymbol{v}) - s_i(\boldsymbol{q})|. \tag{A.13}$$

It is worth noting that in the case of supports with identical cardinality and equidistant

parameter space gridding, the EMD directly represents the average parameter estimation accuracy in gridding units of $\boldsymbol{\Psi}$.

A.2.2 Supports with non-identical cardinalities

When the support sizes are not equal ($K_{\boldsymbol{v}} \neq K_{\boldsymbol{q}}$), the matrix $\boldsymbol{A}$ defined in (A.6) seizes to be a square sub-matrix of $\boldsymbol{F}$ and becomes a rectangular matrix of size $K_{\boldsymbol{v}} \times K_{\boldsymbol{q}}$ instead. In this case, the rows of $\boldsymbol{A}$ sum to $\frac{1}{K_{\boldsymbol{v}}}$ and the columns of $\boldsymbol{A}$ sum to $\frac{1}{K_{\boldsymbol{q}}}$, i.e., $\forall j \in [1, K_{\boldsymbol{q}}]\ \sum_i a_{i,j} = \frac{1}{K_{\boldsymbol{v}}}$ and $\forall i \in [1, K_{\boldsymbol{v}}]\ \sum_j a_{i,j} = \frac{1}{K_{\boldsymbol{q}}}$. Inserting these conditions in (A.5), we have

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \min_{\boldsymbol{A}} (\mathbf{1}_{K_{\boldsymbol{v}}}^{\mathrm{T}} (\boldsymbol{A} \odot \boldsymbol{D}) \mathbf{1}_{K_{\boldsymbol{q}}}) \text{ s.t. } \begin{cases} \boldsymbol{A}\mathbf{1}_{K_{\boldsymbol{q}}} & = \frac{1}{K_{\boldsymbol{v}}} \mathbf{1}_{K_{\boldsymbol{v}}} \\ \boldsymbol{A}^{\mathrm{T}}\mathbf{1}_{K_{\boldsymbol{v}}} & = \frac{1}{K_{\boldsymbol{q}}} \mathbf{1}_{K_{\boldsymbol{q}}} \end{cases}, \tag{A.14}$$

where $\boldsymbol{D}$ is defined the same way as in (A.7).

Due to the different support sizes, (A.14) does not admit a solution where each row and column of $\boldsymbol{A}$ contain a single non-zero element as in (A.8). However, since $\boldsymbol{A}\mathbf{1}_{K_{\boldsymbol{q}}} = \frac{1}{K_{\boldsymbol{v}}}\mathbf{1}_{K_{\boldsymbol{v}}}$ and $\boldsymbol{A}^{\mathrm{T}}\mathbf{1}_{K_{\boldsymbol{v}}} = \frac{1}{K_{\boldsymbol{q}}}\mathbf{1}_{K_{\boldsymbol{q}}}$, we can represent $\boldsymbol{A}$ as $\boldsymbol{A} = \frac{1}{\rho}\bar{\boldsymbol{A}}$ where $\bar{\boldsymbol{A}}$ is a $K_{\boldsymbol{v}} \times K_{\boldsymbol{q}}$ matrix with elements $\bar{a}_{i,j} \in \mathbb{N}_0$ and ρ is the least common multiple of $K_{\boldsymbol{v}}$ and $K_{\boldsymbol{q}}$. In so doing, we obtain

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \frac{1}{\rho} \min_{\bar{\boldsymbol{A}}} (\mathbf{1}_{K_{\boldsymbol{v}}}^{\mathrm{T}} (\bar{\boldsymbol{A}} \odot \boldsymbol{D}) \mathbf{1}_{K_{\boldsymbol{q}}}) \text{ s.t. } \begin{cases} \bar{\boldsymbol{A}}\mathbf{1}_{K_{\boldsymbol{q}}} & = \frac{\rho}{K_{\boldsymbol{v}}} \mathbf{1}_{K_{\boldsymbol{v}}} \\ \bar{\boldsymbol{A}}^{\mathrm{T}}\mathbf{1}_{K_{\boldsymbol{v}}} & = \frac{\rho}{K_{\boldsymbol{q}}} \mathbf{1}_{K_{\boldsymbol{q}}} \end{cases}. \tag{A.15}$$

Now, consider the cost function $R(\bar{\boldsymbol{A}}) = \mathbf{1}_{K_{\boldsymbol{v}}}^{\mathrm{T}} (\bar{\boldsymbol{A}} \odot \boldsymbol{D}) \mathbf{1}_{K_{\boldsymbol{q}}} = \sum_i \sum_j \bar{a}_{i,j} d_{i,j}$. Since all $\bar{a}_{i,j} \in \mathbb{N}_0$, we can represent each non-zero product $\bar{a}_{i,j} d_{i,j}$ in the double sum as an $\bar{a}_{i,j}$-term sum of $d_{i,j}$, i.e., $R(\bar{\boldsymbol{A}}) = \sum_{i,j:\bar{a}_{i,j} \neq 0} \sum_{p=1}^{\bar{a}_{i,j}} d_{i,j}$. This way, the matrix $\bar{\boldsymbol{A}}$ can be expanded into a $\rho \times \rho$ doubly stochastic matrix $\ddot{\boldsymbol{A}}$ and (A.15) can hence be written as

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \frac{1}{\rho} \min_{\ddot{\boldsymbol{A}}} (\mathbf{1}_{\rho}^{\mathrm{T}} (\ddot{\boldsymbol{A}} \odot \ddot{\boldsymbol{D}}) \mathbf{1}_{\rho}) \text{ s.t. } \sum_{\ell=1}^{\rho} \ddot{a}_{\ell,t} = \sum_{t=1}^{\rho} \ddot{a}_{\ell,t} = 1, \tag{A.16}$$

where $\ddot{\boldsymbol{D}} = \boldsymbol{D} \otimes \mathbf{1}_{\frac{\rho}{K_{\boldsymbol{v}}} \times \frac{\rho}{K_{\boldsymbol{q}}}}$ with $\otimes$ being the Kronecker matrix product and $\mathbf{1}_{a \times b}$ denoting an $a \times b$ all-one matrix. Applying the same reasoning as in the case of equally sized supports, we obtain the following (explicit) solution to (A.14)

$$d_{\mathrm{EMD}}(\mathcal{S}(\boldsymbol{v}), \mathcal{S}(\boldsymbol{q})) = \frac{1}{\rho} \sum_{\ell=1}^{\rho} |\ddot{s}_{\ell}(\boldsymbol{v}) - \ddot{s}_{\ell}(\boldsymbol{q})|, \tag{A.17}$$

where $\ddot{s}_{\ell}(\boldsymbol{v})$ and $\ddot{s}_{\ell}(\boldsymbol{q})$ are ℓth elements of the Kronecker-repeated vectors $[s_1(\boldsymbol{v}), \ldots, s_{K_{\boldsymbol{v}}}(\boldsymbol{v})] \otimes \mathbf{1}_{\frac{\rho}{K_{\boldsymbol{v}}}}^{\mathrm{T}}$ and $[s_1(\boldsymbol{q}), \ldots, s_{K_{\boldsymbol{q}}}(\boldsymbol{q})] \otimes \mathbf{1}_{\frac{\rho}{K_{\boldsymbol{q}}}}^{\mathrm{T}}$, respectively.

Appendix B

Proofs and derivations

B.1 Proofs and derivations for Chapter 3

B.1.1 Proof of Lemma 1 in Section 3.3.2.1

Since $Y \sim \chi^2_M$, we can write it is as $Y = \sum_{i=1}^{L} X_i^2$, where X_i are normal random variables, i.e., $X_i \sim \mathcal{N}(0,1)$. Hence $X_i^2 \sim \chi_1^2 \stackrel{\Delta}{=} \Gamma\left(\frac{1}{2}, 2\right)$. Multiplying Y by c, we get

$$cY = \sum_{i=1}^{L} cX_i^2, \tag{B.1}$$

where, due to the scaling property of Gamma distribution

$$cX_i^2 \sim c\Gamma\left(\frac{1}{2}, 2\right) = \Gamma\left(\frac{1}{2}, 2c\right). \tag{B.2}$$

Applying the summation rule, we finally have

$$\sum_{i=1}^{L} X_i^2 \sim \Gamma\left(\frac{L}{2}, 2c\right). \tag{B.3}$$

B.1.2 Proof of Proposition 1 in Section 3.3.2.1

Re-writing (3.15) via $\eta_{\mathrm{O}}^{\mathrm{L}}$, we get

$$\eta_{\mathrm{O}}^{\mathrm{L}} = \vartheta^{\mathrm{L}} + 10\log_{10}(\beta_{1,n}), \tag{B.4}$$

where $\vartheta^{\mathrm{L}} = 10\log_{10}\left(\frac{\|\boldsymbol{x}\|_2^2}{M\sigma_0^2}\right)$ is the ratio of the (instantaneous) signal power to the total noise power expressed in dB. It is convenient to write (B.4) via $\ln(\beta_1)$ as

$$\eta_{\mathrm{O}}^{\mathrm{L}} = \vartheta^{\mathrm{L}} + k_{\mathrm{a}}\ln(\beta_1), \tag{B.5}$$

where $k_{\mathrm{a}} = \frac{10}{\ln(10)}$. As $\beta_1 \sim \Gamma\left(\frac{M}{2}, \frac{2}{M}\right)$, its logarithm is distributed according to the exponential-gamma (EGa) distribution sometimes also referred to as the three-parameter exponential-gamma distribution or generalized Gumbel distribution [169] (see Appendix E.1) with parameters $\mu = \ln\left(\frac{2}{M}\right)$, $\theta = 1$, and $k = \frac{M}{2}$ i.e.,

$$f_{\boldsymbol{A}}(\ln(\beta_1)) = \mathrm{EGa}\left(\ln\left(\frac{2}{M}\right), 1, \frac{M}{2}\right). \tag{B.6}$$

Using the fact that EGa is closed under translation and scaling by a positive factor, we have

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O}}^{\mathrm{L}}) = \mathrm{EGa}\left(k_{\mathrm{a}} \ln\left(\frac{2}{M}\right) + \vartheta_{\mathbb{E}}^{\mathrm{L}}, k_{\mathrm{a}}, \frac{M}{2}\right), \text{ whereas} \tag{B.7}$$

$$f_{\boldsymbol{A}}(\eta_{\mathrm{O/I}}^{\mathrm{L}}) = \mathrm{EGa}\left(k_{\mathrm{a}} \ln\left(\frac{2}{M}\right) + k_{\mathrm{a}} \ln\left(\frac{\gamma}{M}\right), k_{\mathrm{a}}, \frac{M}{2}\right) \tag{B.8}$$

with $\vartheta_{\mathbb{E}}^{\mathrm{L}} = 10 \log_{10}\left(\frac{P_{\mathrm{s}}}{M\sigma_0^2}\right)$.

B.1.3 Proof of Lemma 2 in Section 3.3.2.2

Note that (3.29) is exactly the same as (3.20). Hence, the mean and variance of $(y|\boldsymbol{x})$ are given by $\mathbb{E}(y|\boldsymbol{x}) = k\theta = \|\boldsymbol{x}\|_2^2$ and $\mathrm{var}_{\boldsymbol{A}}\{\eta_{\mathrm{O}}|\mathcal{X}\} = k\theta^2 = \frac{2}{M}\|\boldsymbol{x}\|_2^4 = \frac{2}{M}\mathbb{E}^2\{y|\boldsymbol{x}\}$, respectively. Since $f_{\boldsymbol{A}}(y)$ depends on $\boldsymbol{x}$ only through $\|\boldsymbol{x}\|_2^2$,

$$\mathbb{E}\{y\} = \mathbb{E}_{\boldsymbol{x}}\{\mathbb{E}\{y|\boldsymbol{x}\}\} = \mathbb{E}_{\boldsymbol{x}}\left\{\|\boldsymbol{x}\|_2^2\right\} = P_{\mathrm{s}}. \tag{B.9}$$

We now notice that $(y^2|\boldsymbol{x})$ is distributed according to the generalized Gamma distribution with parameters $p = 0.5$, $d = 0.5k$ and $a = \theta^2$. Taking this into account, we have that $\mathbb{E}\{y^2|\boldsymbol{x}\} = \theta^2 \frac{\Gamma(k+2)}{\Gamma(k)} = \theta^2(k+1)k$ and hence

$$\mathbb{E}\left\{y^2\right\} = \left(\frac{M}{2} + 1\right)\frac{2}{M}P_{\mathrm{s}}^2 = \left(1 + \frac{2}{M}\right)P_{\mathrm{s}}^2. \tag{B.10}$$

The variance of y can now be calculated as

$$\mathrm{var}\{y\} = \mathbb{E}\{y^2\} - \mathbb{E}^2\{y\} = \left(1 + \frac{2}{M}\right)P_{\mathrm{s}}^2 - P_{\mathrm{s}}^2 = \frac{2}{M}P_{\mathrm{s}}^2. \tag{B.11}$$

B.1.4 Derivation of (3.31) and (3.32) in Section 3.3.2.2

In this section, we derive the mean and the variance of $d_m^2 = \left(\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k\right)^2$ where $a_{m,k} \sim \mathcal{B}_e(p)$ and $x_k = x \in \mathbb{R}\ \forall k \in [1, K]$. First, we notice that since all $x_k = x$, we have

$$d_m^2 = x^2\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}\right)^2. \tag{B.12}$$

Now, we denote by $\alpha_\ell = \sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}$ and by $P_\mathrm{s} \triangleq \mathbb{E}\{\boldsymbol{x}\} = K\mathbb{E}\left\{x^2\right\}$. This results in

$$\mathbb{E}\left\{d_m^2\right\} = \mathbb{E}_{\mathcal{X}}\{\mathbb{E}_{\boldsymbol{A}}\{d_m^2|\mathcal{X}\}\} = \mathbb{E}_{\mathcal{X}}\{x^2\mathbb{E}\left\{\alpha_\ell^2\right\}\} = \frac{P_\mathrm{s}}{K}\mathbb{E}\left\{\alpha_\ell^2\right\}, \text{ and} \tag{B.13}$$

$$\mathrm{var}\left\{d_m^2\right\} = \mathbb{E}_{\mathcal{X}}\{x^4\mathbb{E}_{\boldsymbol{A}}\{\alpha_\ell^4\}\} - \frac{P_\mathrm{s}^2}{K^2}\mathbb{E}\left\{\alpha_\ell^2\right\}^2 = \frac{P_\mathrm{s}^2}{K^2}\mathrm{var}\left\{\alpha_\ell^2\right\}, \tag{B.14}$$

From (B.13) and (B.14), to compute the mean and variance of d_m^2, we need to determine the mean and variance of α_ℓ^2. We can calculate the former as

$$\mathbb{E}\left\{\alpha_\ell^2\right\} = \mathrm{var}\left\{\alpha_\ell\right\} + \mathbb{E}\left\{\alpha_\ell\right\}^2 = Kp(1-p) + K^2p^2 = ((K-1)p+1)Kp. \tag{B.15}$$

On the other hand, the variance of α_ℓ^2 can be calculated as

$$\mathrm{var}\left\{\alpha_\ell^2\right\} = \mathbb{E}\left\{\alpha_\ell^4\right\} - \mathbb{E}\left\{\alpha_\ell^2\right\}^2, \tag{B.16}$$

which requires finding the value of $\mathbb{E}\left\{\alpha_\ell^4\right\}$. To do so, we use the fact that α_ℓ is a sum of K i.i.d. Bernoulli random variables and α_ℓ^4 can hence be written as $\alpha_\ell^4 = \left(\sum_{k=1}^K z_k\right)^4$, where $z_k \sim \mathcal{B}_e(p)$. Using multinomial theorem (see App. **??**), we can represent α_ℓ^4 as

$$\alpha_\ell^4 = \sum_{k=1}^K z_k^4 + 4\sum_{i\neq j} z_i^3 z_j + 6\sum_{i\neq j} z_i^2 z_j^2 + 12\sum_{i\neq j\neq q} z_i^2 z_j z_q + 24\sum_{i\neq j\neq q\neq p} z_i z_j z_q z_p. \tag{B.17}$$

Since $z_k \sim \mathcal{B}_e(p)$, we have that $\mathbb{E}\{z_k^n\} = \mathbb{E}\{z_k\} = p$. Furthermore, due to the fact that z_k are independent, the following equalities hold

$$\mathbb{E}\left\{\prod_{1\leq q\leq K} z_q^{k_q}\right\} = \prod_{1\leq q\leq K}\mathbb{E}\left\{z_q^{k_q}\right\} = \prod_{1\leq q\leq K}\mathbb{E}\{z_q\}. \tag{B.18}$$

Applying (B.18) to (B.17), we get

$$\begin{aligned}\mathbb{E}\left\{\alpha_\ell^4\right\} =& \sum_{k=1}^K \mathbb{E}\left\{z_k^4\right\} + 4\sum_{i\neq j}\mathbb{E}\{z_i\}\mathbb{E}\{z_j\} + 6\sum_{i\neq j}\mathbb{E}\{z_i\}\mathbb{E}\{z_j\} \\ &+ 12\sum_{i\neq j\neq q}\mathbb{E}\{z_i\}\mathbb{E}\{z_j\}\mathbb{E}\{z_q\} + 24\sum_{i\neq j\neq q\neq p}\mathbb{E}\{z_i\}\mathbb{E}\{z_j\}\mathbb{E}\{z_q\}\mathbb{E}\{z_p\} \\ =& Kp + 4K(K-1)p^2 + 6\mathrm{C}_2^K p^2 + 12K\mathrm{C}_2^{K-1}p^3 + 24\mathrm{C}_4^K p^4,\end{aligned} \tag{B.19}$$

where $\mathrm{C}_n^K = \dfrac{K!}{n!(K-n)!}$.

Substituting $\mathbb{E}\left\{\alpha_\ell^4\right\}$ and $\mathbb{E}\left\{\alpha_\ell^2\right\}^2$ into (B.16), we obtain

$$\begin{aligned}
&\text{var}\left\{\alpha_\ell^2\right\} = \\
&Kp + 7K(K-1)p^2 + 6K(K-1)(K-2)p^3 + K(K-1)(K-2)(K-3)p^4 - ((K-1)p+1)^2K^2p^2 \\
&= \left(1 + (K-1)(7 + 6(K-2)p + (K^2-5K+6)p^2 - (K-1)^2Kp^3 + 2(K-1)Kp^2 + Kp\right)Kp \\
&= \left(1 - Kp + (K-1)((K^2-5K+6-(K-1)K)p^2 + (4K-12)p + 7)p\right)Kp \\
&= (1-p)p((1 + 2p(K-1)((2K-3)p+3))\,K. \qquad \text{(B.20)}
\end{aligned}$$

Finally, by substituting (B.15) into (B.13) and (B.20) into (B.14), we arrive at

$$\mathbb{E}\left\{d_m^2\right\} = ((K-1)p+1)pP_{\text{s}} \qquad \text{(B.21)}$$

$$\text{var}\left\{d_m^2\right\} = (1-p)p((1+2p(K-1)((2K-3)p+3))\,\frac{P_{\text{s}}}{K}. \qquad \text{(B.22)}$$

B.1.5 Proof of Lemma 3 in Section 3.3.2.2

Since $\beta = \left(\sum_{k=1}^{K} a_k\right)^2$ and $a_k = \pm 1$, we know that β can take the values $(K-2n)^2$ only, where $n = 0, 2, \ldots, \alpha$ and $\alpha = \lfloor 0.5K \rfloor$. Moreover, when $K = 2\alpha + 1$, the set of values of β spans all odd numbers from 1 to α, whereas for $K = 2\alpha$ it contains all even numbers from 0 to α. For the sake of clarity, we will consider these two cases separately.

1) Odd K ($K = 2\alpha + 1$). The probability that among K i.i.d random variables that take the value ± 1 with $p = 0.5$ exactly $n \in [0, \alpha]$ are negative and $K - n$ are positive is $C_n^K p^n p^{K-n} = C_n^K p^K$ since there are C_n^K ways to choose n positions out of K. The same holds for the case of exactly $n \in [0, \alpha]$ of a_k being positive and $K - n$ negative. Now, since we take the square of the summation over a_k, the events when exactly n of a_k are positive and exactly n are negative result in the same values of β. Hence, we have that the PMF of β is given by

$$\Pr\left[\beta = (K-2n)^2\right] = 2C_n^K p^K, \qquad \text{(B.23)}$$

where $n = 0, 1, \cdots \alpha$ and $p = 0.5$.

Expected value: in order to calculate the mean of β we will use the definition of the expected value for a discrete probability distribution, i.e.,

$$\mathbb{E}\left\{\beta\right\} = \sum_{n=0}^{\alpha}(K-2n)^2 \Pr\left[\beta = (K-2n)^2\right] = 2p^K \underbrace{\sum_{n=0}^{\alpha}(K-2n)^2 C_n^K}_{F} = p^{(K-1)}F, \qquad \text{(B.24)}$$

where $F = \sum_{n=0}^{\alpha}(K-2n)^2 C_n^K$ and the last equality is due to the fact that $p^K = \frac{1}{2^K}$. Before proceeding with the calculation of $\mathbb{E}\left\{\beta\right\}$, we introduce a helpful lemma.

Lemma 5. *Let K be a natural number greater than 1 and n be a non-negative integer. Then,*

$$(K-2n)+\frac{1}{n+1}\big(K-2(n+1)\big)^2=\frac{1}{n+1}\big(K-n-1\big)\big(K-2(n+1)\big)+2. \tag{B.25}$$

Proof. The proof proceeds as follows:

$$\begin{aligned}(K-2n)+\frac{1}{n+1}(K-2(n+1))^2 &= (K-n-1)+\frac{1}{n+1}(K-n-1-(n+1))^2-n+1\\ &= (K-n-1)+\frac{1}{n+1}(K-n-1)(K-3n-3)+\frac{1}{n+1}(n+1)^2-n+1\\ &= \frac{1}{n+1}(K-n-1)(n+1+K-3n-3)+2\\ &= \frac{1}{n+1}(K-n-1)(K-2(n+1))+2.\end{aligned}$$

■

Returning to (B.24), we express F in terms of its components as

$$\begin{aligned}F &= K^2+(K-2)^2K+(K-4)^2\frac{K!}{2(K-2)!}+\ldots(K-2\alpha)^2\frac{K!}{\alpha!(K-\alpha)!}\\ &= K\left(K+(K-2)^2+(K-4)^2\frac{(K-1)!}{2(K-2)!}+\cdots+(K-2\alpha)^2\frac{(K-1)!}{\alpha!(K-\alpha)!}\right).\end{aligned} \tag{B.26}$$

Consequently applying Lemma 5 to (B.26), we obtain the following expression for F

$$F=K\Big(2+g(1)\Big(2+g(2)\big(2+\ldots(2+g(\alpha))\big)\Big)\Big)=K\left(2\left(1+\sum_{n=1}^{\alpha-1}\prod_{k=1}^{n}g(k)\right)+\prod_{n=1}^{\alpha}g(n)\right), \tag{B.27}$$

where $g(n)=\frac{1}{n}(K-n), n\in\mathbb{N}$. Since $g(n)=\frac{1}{n}(K-n), n\in\mathbb{N}$, the product of the first n values of $g(n)$ is equal to the binomial coefficient C_n^K, and (B.27) can therefore be written as

$$F=K\left(2\sum_{n=0}^{\alpha-1}C_n^{K-1}+C_\alpha^{K-1}\right)=K\sum_{n=0}^{K-1}C_n^{K-1}=K2^{(K-1)}. \tag{B.28}$$

The fist equality in (B.28) is due to the palindromic property of the binomial coefficients, while the last one is because $\sum_{k=0}^{N}C_k^N=2^N$. Substituting (B.28) into (B.24), we finally obtain

$$\mathbb{E}\{\beta\}=p^{(K-1)}K2^{(K-1)}=K. \tag{B.29}$$

Variance: to compute $\operatorname{var}\{\beta\}$, we make use of the equality $\operatorname{var}\{\beta\}=\mathbb{E}\{\beta^2\}-\mathbb{E}\{\beta\}^2$, which requires finding the expected value of β^2. It can be calculated using the PMF of β^2 as

$$\mathbb{E}\{\beta^2\}=2p^K\underbrace{\sum_{n=0}^{\alpha}(K-2n)^4C_n^K}_{\tilde{F}}. \tag{B.30}$$

Writing $\tilde{F}$ in a way similar to (B.26), i.e.,

$$\tilde{F} = K^4 + (K-2)^4 K + (K-4)^4 \frac{K!}{2(K-2)!} + (K-6)^4 \frac{K!}{6(K-3)!} + \dots (K-2\alpha)^4 \frac{K!}{\alpha!(K-\alpha)!}$$
$$= K\left(K^3 + (K-2)^4 + (K-4)^4 \frac{(K-1)}{2} + \dots + (K-2\alpha)^4 \frac{(K-1)!}{\alpha!(K-\alpha)!}\right). \tag{B.31}$$

Employing the same approach as in (B.26), we arrive at

$$\tilde{F} = K(3K-2)\left(2\sum_{n=0}^{\alpha-1} C_n^{K-1} + C_\alpha^{K-1}\right) = K(3K-2)\sum_{n=0}^{K-1} C_n^{K-1} = K(3K-2)2^{K-1}. \tag{B.32}$$

Substituting (B.32) into (B.30) we obtain

$$\operatorname{var}\{\beta\} = \mathbb{E}\{\beta^2\} - \mathbb{E}\{\beta\}^2 = p^{K-1}K(3K-1)2^{K-1} - K^2 = 2K(K-1). \tag{B.33}$$

2. Even K ($K = 2\alpha$). Clearly, the set of values that β can take for $K = 2\alpha$ stays $\{(K-2n)^2\}_{n=1}^{\alpha}$. Furthermore, the probability that β is equal to $(K-2n)^2$ for $n = 0, 1, \dots, \alpha-1$ in the case of even K is also exactly the same as when K is odd. However, the situation is slightly different when $n = \alpha$ and $K = 2\alpha$. In this case for $n = \alpha$ the value of β is 0, and hence there is only half number of combinations to achieve it (compared to $\sqrt{\beta} = \pm(K-2n)$, for $n = 0, 1, \dots, \alpha-1$). Therefore, the PMF of β is given by

$$\Pr\left[\beta = (K-2n)^2\right] = \begin{cases} 2C_n^K p^K, & n = 0, 1, \cdots \alpha-1 \\ C_n^K p^K, & n = \alpha. \end{cases} \tag{B.34}$$

Expected value: using the fact that $K - 2\alpha = 0$, we write (B.24)

$$\mathbb{E}\{\beta\} = 2p^K \sum_{n=0}^{\alpha} (K-2n)^2 C_n^K = 2p^K \sum_{n=0}^{\alpha-1} (K-2n)^2 C_n^K = p^{(K-1)}F, \tag{B.35}$$

where F now is $\sum_{n=0}^{\alpha-1}(K-2n)^2 C_n^K$. Following the same line of reasoning as above we get

$$F = K\Big(2 + g(1)\Big(2 + g(2)\big(2 + \dots (2 + g(\alpha-1))\big)\Big)\Big) = K\left(2\left(1 + \sum_{n=1}^{\alpha-2}\prod_{k=1}^{n} g(k)\right) + \prod_{n=1}^{\alpha-1} g(n)\right)$$
$$= K\left(2\sum_{n=0}^{\alpha-2} C_n^{K-1} + C_{\alpha-1}^{K-1}\right) = K\sum_{n=0}^{K-1} C_n^{K-1} = K2^{(K-1)}. \tag{B.36}$$

Note that since $K-1$ is odd there is only $2(\alpha-2)+1 = 2\alpha-1$ binomial coefficients C_n^{K-1}. Hence, $\left(2\sum_{n=0}^{\alpha-2} C_n^{K-1} + C_{\alpha-1}^{K-1}\right) = \sum_{n=0}^{K-1} C_n^{K-1}$ and (B.36) we have $\mathbb{E}\{\beta\} = p^{(K-1)}K2^{(K-1)} = K$.

B.1.6 Analysis of Gaussian, Bernoulli and Rademacher random sensing matrices in view of the composite model (3.42) in Section 3.3.3.2

We revisit here three random matrix types considered above in view of the composite structure of $\boldsymbol{A}$. Note that, for simplicity, we consider a real-valued model, while the extension to the complex-valued variables is straightforward as follows from the discussion in Section 3.3.3.1.

Gaussian $\boldsymbol{\Phi}$. When $\phi_{n,m} \sim \mathcal{N}\left(0, \sqrt{1/M}\right)$ each $a_{m,n}$ is a linear combination of N zero-mean normal random variables and hence is a zero-mean random variable itself. However, the elements in each row of $\boldsymbol{A}$ are correlated. This is because for any fixed m all $a_{m,n}$ result from (different) linear combinations of the same random vector $\boldsymbol{\phi}_m$. We know nevertheless that $d_m = \sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k$ is also zero-mean and normal as it is a linear combination of zero-mean normal random variables, albeit correlated [335]. What is left to calculate is the variance of d_m.

Lemma 6. *Let $d_m = \sum_k a_{m,k}x_k$ be a linear combination of K random variables $a_{m,k} = \sum_{n=1}^{N} \phi_{m,n}\psi_{k,n}$ where $x_k \in \mathbb{R}$, $\psi_{k,n} \in \mathbb{R}$ is an (k,n)-th element of $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ and $\phi_{m,n} \sim \mathcal{N}\left(0, \frac{1}{M}\right)$ with $M = 1,\dots M$, $N = 1,\dots N$ and $K \leq N$. The variance of d_m is then given by*

$$\sigma_m^2 = \frac{1}{M}\left(\sum_k x_k^2\epsilon_k^2 + 2\sum_{i,j\in[1,K]:i<j} x_i x_j \left[\boldsymbol{G}_{\boldsymbol{\Psi}}\right]_{i,j}\right), \tag{B.37}$$

where $\epsilon_n = \sqrt{\sum_n \psi_{k,n}^2}$ and $\boldsymbol{G}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}$.

Proof. To calculate the variance of $d_m = \sum_k a_{m,k}x_k$, we first compute the covariance matrix Σ_m of $\boldsymbol{a}_m = [a_{m,1},\dots,a_{N,1}]$:

$$\Sigma_m = \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\phi}}\boldsymbol{\Psi}, \tag{B.38}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}$ denotes the covariance matrix of $\boldsymbol{\phi}_n = [\phi_{n,1},\dots,\phi_{N,1}]$. Noting that $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}$ is a diagonal matrix with $1/M$ on its main diagonal, we can write (B.38) as

$$\Sigma_m = \frac{1}{M}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi} = \frac{1}{M}\boldsymbol{G}_{\boldsymbol{\Psi}}, \tag{B.39}$$

where $\boldsymbol{G}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}$. Finally, denoting by ϵ_n the ℓ_2 norm of the nth column of $\boldsymbol{\Psi}$, the variance of d_m can be expressed as

$$\sigma_m^2 = \frac{1}{M}\left(\sum_k x_k^2\epsilon_k^2 + 2\sum_{i,j\in[1,K]:i<j} x_i x_j \left[\boldsymbol{G}_{\boldsymbol{\Psi}}\right]_{i,j}\right). \tag{B.40}$$

■

Thus, for an orthogonal $\boldsymbol{\Psi}$ we have that $\sigma_m^2 = \frac{\epsilon^2 P_s}{M}$, whereas when it is orthonormal $\sigma_m^2 = \frac{P_s}{M}$ as in the canonical case. Furthermore, when all x_k are equal, the variance of d_m

becomes $\sigma_m^2 = \frac{P_s}{M}\|\boldsymbol{G}_{\boldsymbol{\Psi}}\|_{1,1}$ where $\|\cdot\|_{p,q}$ denotes the $L_{p,q}$ matrix norm. Moreover, since the value of σ_m^2 does not depend on the column index m and the elements of the sensing matrix are i.i.d. within one column, we can invoke Lemma 1 to find the distribution of $\beta = \sum_{m=1}^{M} d_m^2$:

$$f_{\boldsymbol{A}}(\beta|\mathcal{X}) = \Gamma\left(\frac{M}{2}, 2\sigma_m^2\right). \tag{B.41}$$

The rest of the analysis follows along the same lines as in the canonical case keeping in mind the change of the scale parameter from $\frac{2P_s}{M}$ to $2\sigma_m^2$.

Bernoulli and Rademacher $\boldsymbol{\Phi}$. Suppose now that the elements of $\boldsymbol{\Phi}$ are drawn from one of the two discrete distributions considered above. The linear combinations (**??**) create difficulties in determining a precise distribution of the elements of $\boldsymbol{A}$ similar to the ones we have encountered in the canonical model. When N is large enough however we can apply the central limit theorem (CLT) to approximate the distribution of $a_{m,n}$ as normal, i.e.,

$$a_{m,n} \overset{N\uparrow}{\sim} \mathcal{N}\left(\mu_{m,n}, \sigma_{m,n}^2\right), \tag{B.42}$$

where $\mu_{m,n}$ and $\sigma_{m,n}^2$ are the mean and the variance of $\boldsymbol{\phi}_m^{\mathrm{T}}\boldsymbol{\psi}_n$, respectively. They can be calculated as follows

- Bernoulli $\boldsymbol{\Phi}$: $\begin{cases} \mu_{m,n} &= p\sum_{i=1}^{N}\psi_{n,i} \\ \sigma_{m,n}^2 &= p(1-p)\sum_{i=1}^{N}\psi_{n,i}^2 = p(1-p)\|\boldsymbol{\psi}_n\|_2^2 = p(1-p)\epsilon_n \end{cases}$;
- Rademacher $\boldsymbol{\Phi}$: $\begin{cases} \mu_{m,n} &= 0 \\ \sigma_{m,n}^2 &= \sum_{i=1}^{N}\psi_{n,i}^2 = \|\boldsymbol{\psi}_n\|_2^2 = \epsilon_n \end{cases}$.

We observe that in the case of the Rademacher measurement matrix, we arrive at a sensing matrix $\boldsymbol{A}$ with zero-mean Gaussian elements that have (possibly) different variances in different columns. The Bernoulli $\boldsymbol{\Phi}$ also results in a sensing matrix with normal elements of varying (within one row) variance that additionally have different (non-zero) expected values.

In both cases, $d_m = \sum_{k\in\mathcal{S}(\boldsymbol{x})} a_{m,k}x_k$ is a normal random variable: zero-mean for a Radema-cher $\boldsymbol{\Phi}$ and with a (non-zero) mean of $\mu_{\mathrm{d}} = p\sum_{k\in\mathcal{S}(\boldsymbol{x})}\|\boldsymbol{\psi}_k\|_1 x_k$ for a Bernoulli $\boldsymbol{\Phi}$. Applying Lemma 6 we can obtain the variance σ_m^2 of d_m as

$$\sigma_m^2 = \begin{cases} p\sigma_\Psi^2, & \text{for Bernoulli } \boldsymbol{\Phi} \\ \sigma_\Psi^2, & \text{for Rademacher } \boldsymbol{\Phi} \end{cases}, \tag{B.43}$$

where $\sigma_\Psi^2 = \left(\sum_{k\in\mathcal{S}(\boldsymbol{x})} x_k^2\epsilon_k^2 + 2\sum_{i,j\in\mathcal{S}(\boldsymbol{x}):i<j} x_i x_j \left[\boldsymbol{G}_{\boldsymbol{\Psi}}\right]_{i,j}\right)$.

B.1.7 Proof of Theorem 2 in Section 3.4.2.2

Substituting (3.50) into (3.52), we have that

$$\begin{aligned}\Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] &= \Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{a}_i x_i + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_i x_i + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}|\right] \\ &= \Pr\left[|\gamma_{j,i}x_i + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}| \geq |\gamma_{i,i}x_i + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}|\right]. \end{aligned} \tag{B.44}$$

Consider now the condition $|\gamma_{j,i}x_i + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}| \geq |\gamma_{i,i}x_i + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}|$:

$$\begin{aligned} |\gamma_{j,i}x_i + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}| &\geq |\gamma_{i,i}x_i + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}| \\ \left(\gamma_{j,i}x_i + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}\right)^2 - \left(\gamma_{i,i}x_i + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}\right)^2 &\geq 0 \\ \left((\gamma_{j,i}-\gamma_{i,i})x_i + (\boldsymbol{a}_j^{\mathrm{T}} - \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n}\right)\left((\gamma_{j,i}+\gamma_{i,i})x_i + (\boldsymbol{a}_j^{\mathrm{T}} + \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n}\right) &\geq 0. \end{aligned} \tag{B.45}$$

Denote by $\alpha_i = \gamma_{i,i} = \|\boldsymbol{a}_i\|_2^2$. Then, the inequality (B.45) is fulfilled when either

$$\begin{cases} (\boldsymbol{a}_j^{\mathrm{T}} - \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n} & > (\alpha_i - \gamma_{j,i})x_i \\ (\boldsymbol{a}_j^{\mathrm{T}} + \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n} & > -(\alpha_i + \gamma_{j,i})x_i \end{cases} \text{ or } \begin{cases} (\boldsymbol{a}_j^{\mathrm{T}} - \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n} & \leq (\alpha_i - \gamma_{j,i})x_i \\ (\boldsymbol{a}_j^{\mathrm{T}} + \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n} & \leq -(\alpha_i + \gamma_{j,i})x_i \end{cases} \tag{B.46}$$

Consider now $z_- = (\boldsymbol{a}_j^{\mathrm{T}} - \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n}$ and $z_+ = (\boldsymbol{a}_j^{\mathrm{T}} + \boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n}$. Since $\boldsymbol{n}$ is a Gaussian vector whose elements are zero-mean with variance σ_0^2, $z_\mp$ is a zero-mean Gaussian variable with variance $\sigma_\mp^2 = \|\boldsymbol{a}_j \mp \boldsymbol{a}_i\|_2^2\sigma_0^2$. We can write $\sigma_\mp^2$ as

$$\sigma_\mp^2 = \|\boldsymbol{a}_j \mp \boldsymbol{a}_i\|_2^2\sigma_0^2 = (\alpha_i + \alpha_j \mp 2\gamma_{j,i})\sigma_0^2 = (\alpha_i \mp \gamma_{j,i})\left(1 + \frac{\alpha_j \mp \gamma_{j,i}}{\alpha_i \mp \gamma_{j,i}}\right)\sigma_0^2. \tag{B.47}$$

Given (B.47), we define a new variable $\bar{z}_\mp = \dfrac{z_\mp}{\sigma_\mp}$, which is now standard normal distributed, i.e., $\bar{z}_\mp \sim \mathcal{N}(0,1)$. Noting that $\frac{\alpha_i x_i^2}{\sigma_0^2} = \bar{\eta}_{\mathrm{O}}$, we can re-write left-hand part of (B.46) as

$$\begin{aligned}\bar{z}_\mp > \pm\frac{(\alpha_i \mp \gamma_{j,i})x_i}{\sqrt{(\alpha_i \mp \gamma_{j,i})\left(1 + \frac{\alpha_j \mp \gamma_{j,i}}{\alpha_i \mp \gamma_{j,i}}\right)\sigma_0^2}} &= \pm\sqrt{\frac{(\alpha_i \mp \gamma_{j,i})x_i^2}{\sigma_0^2}}\sqrt{\left(1 + \frac{\alpha_j \mp \gamma_{j,i}}{\alpha_i \mp \gamma_{j,i}}\right)^{-1}} \\ &= \pm\sqrt{\bar{\eta}_{\mathrm{O}}}\sqrt{\frac{1 \mp \frac{\gamma_{j,i}}{\alpha_i}}{1 + \frac{\alpha_j \mp \gamma_{j,i}}{\alpha_i \mp \gamma_{j,i}}}}. \end{aligned} \tag{B.48}$$

Similarly, the right-hand part of (B.46) turns into

$$\bar{z}_\mp \leq \pm\sqrt{\frac{1 \mp \frac{\gamma_{j,i}}{\alpha_i}}{1 + \frac{\alpha_j \mp \gamma_{j,i}}{\alpha_i \mp \gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}. \tag{B.49}$$

Finally, combining (B.48) and (B.49) and taking into account that $\bar{z}_{\mp} \sim \mathcal{N}(0,1)$, we obtain

$$\begin{aligned}\Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] &= \Pr\left[\left\{\begin{array}{ll}\bar{z}_- & > \sqrt{\frac{1-\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j-\gamma_{j,i}}{\alpha_i-\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}} \\ \bar{z}_+ & > -\sqrt{\frac{1+\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j+\gamma_{j,i}}{\alpha_i+\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}\end{array}\right. \cup \left\{\begin{array}{ll}\bar{z}_- & \leq \sqrt{\frac{1-\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j-\gamma_{j,i}}{\alpha_i-\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}} \\ \bar{z}_+ & \leq -\sqrt{\frac{1+\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j+\gamma_{j,i}}{\alpha_i+\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}\end{array}\right.\right] \\ &= Q\left(\sqrt{\frac{1-\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j-\gamma_{j,i}}{\alpha_i-\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}\right)\cdot Q\left(-\sqrt{\frac{1+\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j+\gamma_{j,i}}{\alpha_i+\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}\right) \\ &+ Q\left(-\sqrt{\frac{1-\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j-\gamma_{j,i}}{\alpha_i-\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}\right)\cdot Q\left(\sqrt{\frac{1+\frac{\gamma_{j,i}}{\alpha_i}}{1+\frac{\alpha_j+\gamma_{j,i}}{\alpha_i+\gamma_{j,i}}}\bar{\eta}_{\mathrm{O}}}\right). \qquad \text{(B.50)}\end{aligned}$$

B.1.8 Proof of Theorem 3 in Section 3.4.2.2

To obtain (3.59), we proceed in a manner similar to the proof of Theorem 2. Namely, we substitute (3.56) into $\Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right]$ to obtain

$$\begin{aligned}\Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] &= \Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\sum_{k\in\mathcal{S}(\boldsymbol{x})}\boldsymbol{a}_k x_k + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}| \geq |\boldsymbol{a}_i^{\mathrm{T}}\sum_{k\in\mathcal{S}(\boldsymbol{x})}\boldsymbol{a}_k x_k + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}|\right] \\ &= \Pr\left[|\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{j,k}x_k + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}| \geq |\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}x_k + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}|\right]. \qquad \text{(B.51)}\end{aligned}$$

Note that $i \in \mathcal{S}(\boldsymbol{x})$ here, i.e., $\exists k \in \mathcal{S}(\boldsymbol{x}) : k = i$.

The condition in (B.51) is equivalent to

$$\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{j,k}x_k + \boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{n}\right)^2 - \left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}x_k + \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{n}\right)^2 \geq 0 \qquad \text{(B.52)}$$

$$\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}-\gamma_{i,k}\right)x_k + \left(\boldsymbol{a}_j^{\mathrm{T}}-\boldsymbol{a}_i^{\mathrm{T}}\right)\boldsymbol{n}\right)\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}+\gamma_{i,k}\right)x_k + \left(\boldsymbol{a}_j^{\mathrm{T}}+\boldsymbol{a}_i^{\mathrm{T}}\right)\boldsymbol{n}\right) \geq 0.$$

Note that (B.52) and (B.46) are similar but for the change of the first term in both brackets. Hence, we can formulate the conditions on $\left(\boldsymbol{a}_j^{\mathrm{T}}-\boldsymbol{a}_i^{\mathrm{T}}\right)\boldsymbol{n}$ and $\left(\boldsymbol{a}_j^{\mathrm{T}}+\boldsymbol{a}_i^{\mathrm{T}}\right)\boldsymbol{n}$ according to (B.46) keeping in mind the change from $\pm(\alpha_i \mp \gamma_{j,i})x_i$ to $\pm\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{i,k}\mp\gamma_{j,k}\right)x_k\right)$.

Taking into account that the random variables $z_- = (\boldsymbol{a}_j^{\mathrm{T}}-\boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n}$ and $z_+ = (\boldsymbol{a}_j^{\mathrm{T}}+\boldsymbol{a}_i^{\mathrm{T}})\boldsymbol{n}$ stay zero-mean Gaussian with variance $\sigma_{\mp}^2 = \|\boldsymbol{a}_j \mp \boldsymbol{a}_i\|_2^2\sigma_0^2 = (\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2$, we can

follow (B.48) and (B.49)to obtain

$$\bar{z}_{\mp} > \pm\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{i,k}\mp\gamma_{j,k}\right)x_k\right)}{\sqrt{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}} = \pm\sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{i,k}\mp\gamma_{j,k}\right)x_k\right)^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}} \text{ or} \tag{B.53}$$

$$\bar{z}_{\mp} \leq \pm\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{i,k}\mp\gamma_{j,k}\right)x_k\right)}{\sqrt{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}} = \pm\sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{i,k}\mp\gamma_{j,k}\right)x_k\right)^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}}, \tag{B.54}$$

where $\bar{z}_{\mp}\sim\mathcal{N}(0,1)$. Denoting by $\alpha_{i,j}^{\mp}=\sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{i,k}\mp\gamma_{j,k}\right)x_k\right)^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}}$ we arrive at

$$\Pr\left[|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{y}|\geq|\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}|\right] = Q\left(\alpha_{j,i}^{-}\right)\cdot Q\left(-\alpha_{j,i}^{+}\right)+Q\left(-\alpha_{j,i}^{-}\right)\cdot Q\left(\alpha_{j,i}^{+}\right). \tag{B.55}$$

B.1.9 Proof of Corollary 1 in Section 3.4.2.2

We begin the proof by re-writing the denominator of $\bar{\eta}_{\mathrm{O}}=\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\sigma_0^2}$ as

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{x}\|_2^2 &= \left\|\sum_{k\in\mathcal{S}(\boldsymbol{x})}\boldsymbol{a}_k x_k\right\|_2^2 = \left\|\sum_{k\in\mathcal{S}(\boldsymbol{x})}\boldsymbol{a}_k\right\|_2^2 x^2 = \left(\sum_{n=1}^{K}\|\boldsymbol{a}_{k_n}\|_2^2+2\sum_{n=1}^{K}\sum_{\ell=1}^{n-1}\boldsymbol{a}_{k_k}^{\mathrm{T}}\boldsymbol{a}_{k_\ell}\right)x^2\\ &= \left(\sum_{n=1}^{K}\gamma_{k_n k_n}+2\sum_{n=1}^{K}\sum_{\ell=1}^{n-1}\gamma_{k_n,k_\ell}\right)x^2 = \sum_{n=1}^{K}\sum_{\ell=1}^{K}\gamma_{k_n,k_\ell}x^2,\end{aligned} \tag{B.56}$$

where $k_n,k_\ell\in\mathcal{S}(\boldsymbol{x})$.

Denote by $\bar{\eta}_{\mathrm{O}}^{(i)}=\frac{\sum_{\ell=1}^{K}\gamma_{k_i,k_\ell}x^2}{\sigma_0^2}$, then $\bar{\eta}_{\mathrm{O}}=\sum_{i=1}^{K}\bar{\eta}_{\mathrm{O}}^{(i)}$. Consider now $\alpha_{i,j}^{\mp}=\sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)x_k\right)^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}}$:

$$\begin{aligned}\alpha_{i,j}^{\mp} &= \sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)x_k\right)^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}} = \sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)\right)^2x^2}{(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}}\\ &= \sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)\right)^2\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}x^2}{\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}\cdot(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})\sigma_0^2}} = \sqrt{\frac{\left(\sum_{k\in\mathcal{S}(\boldsymbol{x})}\left(\gamma_{j,k}\mp\gamma_{i,k}\right)\right)^2}{\sum_{k\in\mathcal{S}(\boldsymbol{x})}\gamma_{i,k}\cdot(\gamma_{i,i}+\gamma_{j,j}\mp 2\gamma_{j,i})}\bar{\eta}_{\mathrm{O}}^{(i)}}.\end{aligned} \tag{B.57}$$

B.2 Proofs and derivations for Chapter 4

B.2.1 Proof of Theorem 4 in Section 4.2.1

Since $\boldsymbol{X}$ is row K-sparse, we can express $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} = \tilde{\boldsymbol{A}}\bar{\boldsymbol{X}}, \tag{B.58}$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{A}_{\mathcal{S}(\boldsymbol{X})} \in \mathbb{C}^{M\times K}$ and $\bar{\boldsymbol{X}} = \boldsymbol{X}^{\mathcal{S}(\boldsymbol{X})} \in \mathbb{C}^{K\times T}$ with $\boldsymbol{X}^{\mathcal{S}(\boldsymbol{X})}$ denoting the rows of $\boldsymbol{X}$ indexed by $\mathcal{S}(\boldsymbol{X})$. On one hand, we know that

$$\operatorname{rank}\left(\tilde{\boldsymbol{A}}\bar{\boldsymbol{X}}\right) \leq \min\left(\operatorname{rank}\left(\tilde{\boldsymbol{A}}\right), \operatorname{rank}\left(\bar{\boldsymbol{X}}\right)\right) = \min(M, K, T) = K. \tag{B.59}$$

On the other hand, from the Sylvester's rank inequality we have that

$$\operatorname{rank}\left(\tilde{\boldsymbol{A}}\right) + \operatorname{rank}\left(\bar{\boldsymbol{X}}\right) - K \leq \operatorname{rank}\left(\tilde{\boldsymbol{A}}\bar{\boldsymbol{X}}\right). \tag{B.60}$$

Since $\operatorname{krank}(\boldsymbol{A}) \geq K$ and $M > K$, $\operatorname{rank}\left(\tilde{\boldsymbol{A}}\right) = K$ and hence

$$K + K - K \leq \operatorname{rank}(\boldsymbol{Y}) = \operatorname{rank}\left(\tilde{\boldsymbol{A}}\bar{\boldsymbol{X}}\right) \leq K. \tag{B.61}$$

From (B.61) it follows that $\operatorname{rank}(\boldsymbol{Y}) = K$, which concludes the proof.

B.2.2 Proof of Theorem 5 in Section 4.4.1

The "only-if" part becomes evident by considering $K = 1$. In this case, we have $\boldsymbol{y}_b = \boldsymbol{A}_b\boldsymbol{x} = \boldsymbol{a}_{b,n}x_n$, where x_n is the value of the single non-zero in $\boldsymbol{x}$ and $\boldsymbol{a}_{b,n}$ is the nth row of $\boldsymbol{A}_b$. For $\boldsymbol{Y}$ we then have $\boldsymbol{Y} = [\boldsymbol{a}_{1,n}, \ldots, \boldsymbol{a}_{B,n}]\, x_n$ which is rank-one only if all columns $\boldsymbol{a}_{1,n}$ are scaled version of one common non-zero vector $\boldsymbol{a}_{0,n}$, i.e., $\boldsymbol{a}_{b,n} = c_{b,n}\boldsymbol{a}_{0,n}$. Stacking $\boldsymbol{a}_{b,n}$ back into $\boldsymbol{A}$ we obtain $\boldsymbol{a}_n = \boldsymbol{c}_n \otimes \boldsymbol{a}_{0,n}$ and therefore $\boldsymbol{A} = \boldsymbol{C} \diamond \boldsymbol{A}_0$.

For the "if" part, consider $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{C}\diamond\boldsymbol{A}_0)\boldsymbol{x}$. Its reshaped version $\boldsymbol{Y}$ can be expressed as $\boldsymbol{Y} = \boldsymbol{A}_0\operatorname{diag}\{\boldsymbol{x}\}\boldsymbol{C}^{\mathrm{T}} = \boldsymbol{A}_{0,K}\operatorname{diag}\{\boldsymbol{x}_K\}\boldsymbol{C}_K^{\mathrm{T}}$, where $\boldsymbol{\Phi}_{0,K} \in \mathbb{C}^{m\times K}$, $\boldsymbol{C}_K \in \mathbb{C}^{B\times K}$ and $\boldsymbol{x}_K$ contain only the K columns/values corresponding to the non-zero entries in $\boldsymbol{x}$, respectively. Since $\boldsymbol{A}_{0,K}$ and $\boldsymbol{C}_K$ have full column rank K and provide a rank factorization of $\boldsymbol{Y}$, we have that $\operatorname{rank}\{\boldsymbol{Y}\} = K$. Obviously, this can only be fulfilled if $K \leq B$ and $K \leq m$.

B.2.3 Proof of Theorem 6 in Section 4.4.1

The proof proceeds in a similar fashion to the proof of Theorem 5. We begin by considering $K = 1$ for the "only-if" part. As before we obtain $\boldsymbol{Y} = [\boldsymbol{a}_{1,n}, \ldots, \boldsymbol{a}_{B,n}]\, x_n$, where $\boldsymbol{a}_{b,n}$ is the b-th block of the n-th column of $\boldsymbol{A}$. To obtain the desired property that $\operatorname{rank}\{\boldsymbol{Y}\} = K = 1$, we therefore require that all $\boldsymbol{a}_{b,n}$ are linearly dependent, i.e., $\boldsymbol{a}_{b,n} = c_{b,n}\boldsymbol{a}_{b-1,n}$. For non-

overlapping blocks, all the $c_{b,n}$ are different, which immediately yields the Khatri-Rao structure with an arbitrary matrix $\boldsymbol{C}$. We now show that for overlapping blocks, there must be much more structure in the coefficients $\boldsymbol{C}$.

Consider three consecutive columns $b-1$, b, and $b+1$. We have the two scaling conditions

$$\boldsymbol{a}_{b,n} = c_{b,n}\boldsymbol{a}_{b-1,n} \tag{B.62}$$

$$\boldsymbol{a}_{b+1,n} = c_{b+1,n}\boldsymbol{a}_{b,n} \tag{B.63}$$

which must be valid for all elements in the vectors. However, since the blocks overlap by $m-p$ samples, they have common elements. In particular,

$$[\boldsymbol{a}_{b-1,n}]_{(k)} = [\boldsymbol{a}_{b,n}]_{(k-p)} \tag{B.64}$$

$$[\boldsymbol{a}_{b,n}]_{(k)} = [\boldsymbol{a}_{b+1,n}]_{(k-p)} \tag{B.65}$$

for all $k = p+1, p+2, \ldots, m$. Inserting (B.62) and (B.63) into (B.65) we obtain

$$c_{b,n}[\boldsymbol{a}_{b-1,n}]_{(k)} = c_{b+1,n}[\boldsymbol{a}_{b,n}]_{(k-p)}. \tag{B.66}$$

However, since (B.64) must be true for all $k = p+1, p+2, \ldots, m$ this implies that $c_{b+1,n} = c_{b,n}$. Since b is arbitrary, the same argument can be applied to show that $c_{b,n} = c_n \; \forall b = 1, 2, \ldots, B$. Therefore, we have

$$\boldsymbol{a}_{b,n} = c_n\boldsymbol{a}_{b-1,n} = c_n^2\boldsymbol{a}_{b-2,n} = \ldots = c_n^{b-1}\boldsymbol{a}_{1,n}. \tag{B.67}$$

Due to the overlap, to satisfy (B.67) we can choose only the first p elements of $\boldsymbol{a}_{1,n}$ as well as the constant $c_n \in \mathbb{C}$ freely. The remaining elements of $\boldsymbol{a}_n$ are filled by replicating these p elements, scaled by c_n, c_n^2, and so on. This leads to the Vandermonde structure of $\boldsymbol{C}$.

For the "if" part, it is easy to see that for every K we can write $\boldsymbol{Y}$ as $\tilde{\boldsymbol{A}}_0 \mathrm{diag}\{\boldsymbol{x}\}\boldsymbol{C}_B^{\mathrm{T}}$, where $\tilde{\boldsymbol{A}}_0$ contains the first m rows of $\boldsymbol{C}_{\left\lceil \frac{m}{p} \right\rceil} \diamond \boldsymbol{A}_0$, and $\boldsymbol{C}_{\left\lceil \frac{m}{p} \right\rceil}$ and $\boldsymbol{C}_B$ represent the first $\left\lceil \frac{m}{p} \right\rceil$ and the first B rows of $\boldsymbol{C}$, respectively. Therefore, we can write $\boldsymbol{Y}$ as $\tilde{\boldsymbol{A}}_{0,K} \mathrm{diag}\{\boldsymbol{x}_K\} \cdot \boldsymbol{C}_{B,K}^{\mathrm{T}}$, where $\tilde{\boldsymbol{A}}_{0,K} \in \mathbb{C}^{m \times K}$, $\boldsymbol{C}_{B,K} \in \mathbb{C}^{B \times K}$, and $\boldsymbol{x}_K$ contain the K columns/values corresponding to the nonzero entries in $\boldsymbol{x}$. Since $\boldsymbol{C}$ and $\boldsymbol{A}_0$ have full Kruskal-rank K, the same holds true for $\tilde{\boldsymbol{A}}_0$ and $\boldsymbol{C}_B$ [336] and thus we have $\mathrm{rank}\{\boldsymbol{Y}\} = K$.

B.3 Proofs and derivations for Chapter 5

B.3.1 Derivation of P_{d}, P_{fa} and P_{md} from Example 5.19 in Section 5.2.2.2

Probability of correct detection P_{d}. According to (5.9), the correct detection occurs when all of the elements of the binary occupancy vector $\boldsymbol{b}$ are estimated correctly. In case of b_n being i.i.d. Bernoulli random variables with $p = 0.5$, there are $2^2 = 4$ possible outcomes but only 3 of them are different from the detection point of view, i.e., a vector that contains all zeros, all ones,a single zero and a single one. Clearly, there is only 1 realization of the first two, while the last two can appear in 2 different variations as demonstrated below.

$b_{q,n} \backslash q$	1	2	2	3
$b_{q,1}$	0	0	1	1
$b_{q,2}$	0	1	0	1

Taking into account that the probability of correctly estimating 1 or 0 is $1-\alpha$ or $1-\beta$, respectively, we can write P_{d} as

$$\begin{aligned} P_{\mathrm{d}} &= \frac{1}{4}\left((1-\alpha)^2 + 2(1-\alpha)(1-\beta) + (1-\beta)^2\right) \\ &= \frac{1}{4}(1-\alpha+1-\beta)^2 = \frac{1}{4}(2-\alpha-\beta)^2. \end{aligned} \tag{B.68}$$

Probability of false alarm P_{fa}. To compute the total probability of false alarm we need to consider all combinations where there is at least a single 0 in the occupancy vector $\boldsymbol{b}$, which is obviously all of them but the vector $[1,1]^{\mathrm{T}}$. Among the rest of possible combinations it is enough to consider only two, i.e., $[0,0]^{\mathrm{T}}$ and $[0,1]^{\mathrm{T}}$. As a result we obtain

$$P_{\mathrm{fa}} = \frac{1}{4}\left(\alpha^2 + 2\alpha(1-\alpha) + 2\alpha\right) = \frac{1}{4}\alpha(4-\alpha). \tag{B.69}$$

.

Probability of false alarm P_{md}. Computation of the total probability of missed detection follows along the same lines that that of P_{fa}, but the change from a factor of α in front of the first bracket to the factor of β. Therefore,

$$P_{\mathrm{md}} = \frac{1}{4}\beta(4-\beta). \tag{B.70}$$

B.4 Proofs and derivations for Chapter 6

B.4.1 Proof of Theorem 7 in Section 6.3

In [10], it is shown that the sparse recovery problem (6.13) has a unique solution if $\text{spark}(\boldsymbol{B}) \geq 2K+1$. In order to proof that this requirement is fulfilled for any fixed $\boldsymbol{a}(\theta, f) \in \mathcal{A}$ and M_t within the limits of the theorem, in the following we show that it is always possible to achieve matrix $\boldsymbol{B}$ such that any of its $r \leq 2K$-column sub-matrices $\boldsymbol{B}_r$ are full rank. We begin with a following lemma.

Lemma 7 (Lemma 1 in [307])**.** *Let $\boldsymbol{A} \in \mathbb{C}^{N\times K}$ and $\boldsymbol{B} \in \mathbb{C}^{M\times K}$ be two matrices with Kruskal rank $\kappa_{\boldsymbol{A}}$ and $\kappa_{\boldsymbol{B}}$, respectively. Then, their Khatri-Rao (column-wise Kronecker) product $\boldsymbol{C} = \boldsymbol{B} \diamond \boldsymbol{A} \in \mathbb{C}^{NM\times K}$ is full rank if*

$$\begin{cases} \text{rank}(\boldsymbol{A}) + \kappa_{\boldsymbol{B}} \geq K+1 \\ \kappa_{\boldsymbol{A}} \geq 1 \end{cases}, \text{ or } \begin{cases} \text{rank}(\boldsymbol{B}) + \kappa_{\boldsymbol{A}} \geq K+1 \\ \kappa_{\boldsymbol{B}} \geq 1 \end{cases}. \tag{B.71}$$

Proof. Proof see Appendix A in [307]. ■

Corollary 3. *If $\boldsymbol{A}$ is full Kruskal rank with $\kappa_{\boldsymbol{A}} = \min(N, K)$ and $\boldsymbol{B}$ does not have zero columns, then $\boldsymbol{C} = \boldsymbol{B} \diamond \boldsymbol{A}$ is full rank when*

$$\text{rank}(\boldsymbol{B}) \geq \begin{cases} 1, & \text{if } N \geq K \\ K - N + 1, & \text{otherwise.} \end{cases} \tag{B.72}$$

Proof. The proof follows directly from Lemma 7. ■

Now, suppose that $\boldsymbol{\Phi}_n = \boldsymbol{\Phi}$ for any $n = 1, 2, \ldots, N$ and consider an $r \leq 2K$-column sub-matrix of $\boldsymbol{B}$ corresponding to some active triples $(\alpha_j, \beta_j, \ell_j)$ with $j = 1, 2, \ldots, r$. It can be written as

$$\boldsymbol{B}_r = (\boldsymbol{1}_{N\times 1} \otimes \boldsymbol{\Phi}_r) \odot (\boldsymbol{a}_r \otimes \boldsymbol{1}_{M\times 1}) = \boldsymbol{\Phi}_r \diamond \boldsymbol{a}_r, \tag{B.73}$$

where $\odot$ stands for the Schur-Hadamard (element-wise) matrix product, $\boldsymbol{a}_r \in \mathbb{C}^{N\times r}$ contains the array steering vectors of the r active pairs $(\alpha_j\theta_{\text{d}}, \beta_j f_{\text{d}})$, and $\boldsymbol{\Phi}_r \in \mathbb{C}^{M\times r}$ contains the columns of the matrix $\boldsymbol{\Phi}$ corresponding to the active spectral cells ℓ_r. Since $\boldsymbol{a}(\theta, f) \in \mathcal{A}$, $\boldsymbol{a}_r$ is full Kruskal rank, i.e., $\kappa_{\boldsymbol{a}_r} = \min(N, r)$. Invoking Corollary 1, the matrix $\boldsymbol{B}_r$ is full rank if

$$\text{rank}(\boldsymbol{\Phi}_r) \geq \begin{cases} 1, & \text{if } N \geq r \\ r - N + 1, & \text{otherwise} \end{cases} \tag{B.74}$$

That is, if $N \geq 2K$ any matrix $\boldsymbol{\Phi}$ with non-zero columns (even with $M = 1$) allows to achieve a full column rank matrix $\boldsymbol{B}_r$.

Consider now the case when $N < r$ and suppose we choose $\boldsymbol{\Phi}$ such that it has full Kruskal rank, i.e., $\kappa_{\boldsymbol{\Phi}} = M$. Invoking the same argument as before, $\boldsymbol{a}_r$ is always full Kruskal rank, i.e., $\kappa_{\boldsymbol{a}_r} = N$. However, if some of the indices $\ell_j \in [1, r]$ coincide, then $\boldsymbol{\Phi}_r$ has some repeated columns and is hence rank deficient. The rank of $\boldsymbol{\Phi}_r$ then becomes

$$\operatorname{rank}(\boldsymbol{\Phi}_r) = \min(M, \mu). \tag{B.75}$$

Substituting (B.75) into (B.74), we have that

$$\begin{cases} MN \geq r \\ N \geq r - \min(M, \mu) + 1 \end{cases}, \tag{B.76}$$

where the first inequality follows from the requirement that for $\boldsymbol{B}_R$ to be full column rank it has to be tall or square. When r takes its maximum value $2K$, we get

$$\begin{cases} MN \geq 2K \\ N \geq 2K - \min(M, \mu) + 1 \end{cases}. \tag{B.77}$$

B.5 Proofs and derivations for Chapter 7

B.5.1 Proof of Lemma 4 in Section 7.3.2.2

Suppose that $\mathcal{S}_{p_1} \neq \mathcal{S}_{p_2}$. Then, there exists such $i \in [1, N]$ that $i \notin \mathcal{C}_{p_1}^{(o)} \cap \mathcal{C}_{p_2}^{(o)}$. Therefore, $\mathcal{C}_{p_1}^{(o)} \triangle \mathcal{C}_{p_2}^{(o)} \neq \emptyset$, where $\triangle$ denotes the symmetric difference between two sets defined as

$$\mathcal{C}_{p_1}^{\mathrm{o}} \triangle \mathcal{C}_{p_2}^{\mathrm{o}} = (\mathcal{C}_{p_1}^{\mathrm{o}} \setminus \mathcal{C}_{p_2}^{\mathrm{o}}) \cup (\mathcal{C}_{p_2}^{\mathrm{o}} \setminus \mathcal{C}_{p_1}^{\mathrm{o}}). \tag{B.78}$$

Invoking the definition of $\mathcal{C}_{p}^{(o)}$, we have that since $\mathcal{C}_{p_1}^{(o)} \triangle \mathcal{C}_{p_2}^{(o)} \neq \emptyset$ there must exist such $k \in [1, K]$ that $f_{\mathrm{c},k} = f_i^{\mathrm{c}}$ and $k \notin \mathcal{K}_{p_1}^{\mathrm{o}} \cap \mathcal{K}_{p_2}^{\mathrm{o}}$. Therefore, the symmetric difference $\mathcal{K}_{p_2}^{\mathrm{o}} \triangle \mathcal{K}^{\mathrm{o}}$ is also a non-empty set and hence $\mathcal{K}_{p_1}^{\mathrm{o}} \neq \mathcal{K}_{p_2}^{\mathrm{o}}$.

B.5.2 Proof of Theorem 9 in Section 7.3.2.2

Necessary condition ('If' part): If $\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2} \neq \emptyset$ and $\mathcal{S}_{p_1,p_2} \neq \mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}$, then there exist ℓ_i such that $\ell_i \in (\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}) \setminus \mathcal{S}_{p_1,p_2}$ where $\setminus$ denotes the relative compliment. Since $(\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}) \setminus \mathcal{S}_{p_1,p_2} \neq \emptyset$ and $\mathcal{S}_{p_1,p_2} \subseteq (\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2})$, there must exist a pair $(k_1 \in \mathcal{K}_{p_1}^{\mathrm{o}}, k_2 \in \mathcal{K}_{p_2}^{\mathrm{o}}), k_1 \neq k_2$ such that $\{k_1, k_2\} \notin \mathcal{K}_{p_1,p_2}^{\mathrm{o}}$ and $f_{\mathrm{c},k_1} = \ell_i f_{\mathrm{p}} = f_{\mathrm{c},k_2}$. Therefore, the set $\mathcal{I}$ is non-empty.

Sufficient condition ('Only if' part): If $\mathcal{I} \neq \emptyset$, then there exist q pair of indices $(k_1 \in [1, K], k_2 \in [1, K]), k_1 \neq k_2$ such that $f_{\mathrm{c},k_1} = f_{\mathrm{c},k_2}$. Since each sensor observes $K_{\mathrm{obs}}^{(n)}$ signals $x_k(t)$ with distinct central frequencies $f_{\mathrm{c},k}$, there is no such sensor that observes both signals $x_{k_1}(t)$ and $x_{k_2}(t)$, i.e., $\nexists k \in [1, K] : \{k_1, k_2\} \in \mathcal{K}_p^{\mathrm{o}}$. However, taking into account that each $x_k(t)$ is observed by at least one sensor, we have that there exists a pair of sensors with indices $(p_1 \in [1, P], p_2 \in [1, L]), p_1 \neq p_2$ such that $k_1 \in \mathcal{K}_{p_1}^{\mathrm{o}}$, $k_2 \in \mathcal{K}_{p_2}^{\mathrm{o}}$ and $\{k_1, k_2\} \notin \mathcal{K}_{p_1,p_2}^{\mathrm{o}}$. However, it also holds that

$$\exists \ell_i \in \mathcal{S}_{p_1} \cap \mathcal{S}_{p_2} : \ell_i \notin \mathcal{S}_{p_1,p_2}, \tag{B.79}$$

where $\ell_i = \left\lfloor \frac{f_{\mathrm{c},k_1}}{f_{\mathrm{p}}} \right\rfloor = \left\lfloor \frac{f_{\mathrm{c},k_2}}{f_{\mathrm{p}}} \right\rfloor$. Therefore, $\mathcal{S}_{p_1} \cap \mathcal{S}_{p_2} \neq \emptyset$ and $\mathcal{S}_{p_1,p_2} \neq \mathcal{S}_{p_1} \cap \mathcal{S}_{p_2}$.

Appendix C

Elements of linear algebra

C.1 Vector and matrix norms

An $\ell_{p\geq 1}$ norm of an length-N vector $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\|_p = \left(\sum_{n=1}^{N} |x_n|^p\right)^{\frac{1}{p}}. \tag{C.1}$$

For $p = 1$ an ℓ_1-norm is called a Taxicab or a Manhattan norm while for $p = 2$, one obtains an ℓ_2-norm or Euclidean norm. The ℓ_p-norm with $p = \infty$ is called the infinity or maximum norm and it is given by

$$\|\boldsymbol{x}\|_\infty = \max_n |x_n|. \tag{C.2}$$

Finally, one also defines an ℓ_0(pseudo)-norm as

$$\|\boldsymbol{x}\|_0 = \sum_{i:x_i\neq 0} 1. \tag{C.3}$$

Entry-wise L_p matrix norms treat an $M \times N$ matrix $\boldsymbol{A}$ as a vector of size MN and apply one of the ℓ_p vector norms introduced above such that

$$\|\boldsymbol{A}\|_p = \|\mathrm{vec}\,(\boldsymbol{A})\,\|_p = \left(\sum_{m=1}^{M}\sum_{n=1}^{N} |a_{m,n}|^p\right)^{1/p}. \tag{C.4}$$

One can also introduce mixed $L_{p,q}$-norm where

$$\|\boldsymbol{A}\|_{p,q} = \left(\sum_{n=1}^{N}\left(\sum_{m=1}^{M} |a_{m,n}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}} \tag{C.5}$$

and $p, q \geq 1$. For $p = q = 2$ the $L_{p,q}$-norm is called a Frobenius norm denoted as $\|\boldsymbol{A}\|_{\mathrm{F}}$ and

given by

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{\sum_{n=1}^{N}\sum_{m=1}^{M}|a_{m,n}|^2} = \sqrt{\mathrm{trace}\,(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}. \tag{C.6}$$

Finally, $p = q = \infty$ yields a max matrix norm, i.e.,

$$\|\boldsymbol{A}\|_{\max} = \max_{m,n}|a_{m,n}|. \tag{C.7}$$

For vector induced matrix norms see [337].

C.2 List of some special matrices

- **All-ones matrix**
  An all-ones matrix is an $M \times N$ matrix where every element is equal to one. It is denoted by $\mathbf{1}_{M\times N}$,
  $$\mathbf{1}_{M\times N} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}.$$
  An N-length vector of all-ones is then simply denoted by $\mathbf{1}_N$.
- **All-zero matrix**
  An all-zeros matrix is a counterpart of an all-ones matrix as all its elements are equal to zero. Similarly to an all-one matrix, an all-zeros matrix of size $M \times N$ is
  $$\mathbf{0}_{M\times N} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix},$$
  while the all-zeros vector of length N is denoted by $\mathbf{0}_N$.
- **Diagonal matrix**
  A diagonal matrix is a square $N \times N$ matrix whose elements are non-identically zero only on the main diagonal, i.e.,
  $$\boldsymbol{A} = \begin{pmatrix} a_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{N,N} \end{pmatrix},$$
  where $a_{i,j} = 0\ \forall i \neq j \in \mathbb{N}_{1:N}$. The operation of obtaining an length-N vector containing the diagonal elements of $\boldsymbol{A}$ is denoted by $\mathrm{diag}\,(\boldsymbol{A})$, e.g., $\boldsymbol{a} = \mathrm{diag}\,(\boldsymbol{A}) = [a_{1,1}, \ldots, a_{N,N}]^{\mathrm{T}}$.

- **A doubly stochastic matrix**
  A doubly stochastic matrix is a non-negative real matrix whose every row and column adds to one, i.e., and $M \times N$ matrix $\boldsymbol{A}$ is doubly stochastic if and only if $\sum_j a_{i,j} = \sum_i a_{i,j} = 1$ for any $i \in [1, M]$ and $j \in [1, N]$.

- **Block diagonal matrix**
  A block diagonal matrix is a block matrix that has $M_n \times N_n$ matrices $\boldsymbol{A}_n$ on its main diagonal, i.e.,
  $$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \cdots & \mathbf{0}_{M_1 \times N_N} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{M_N \times N_1} & \cdots & \boldsymbol{A}_N \end{pmatrix}.$$
  The size of $\boldsymbol{A}$ is $\sum_{n=1}^{N} M_n \times \sum_{n=1}^{N} N_n$.

- **Identity matrix**
  An identity matrix is a diagonal matrix that has ones on its main diagonal and it is denoted by $\mathbf{I}_N$, i.e.
  $$\mathbf{I}_N = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}.$$

- **Gram matrix**
  Gram or Gramian matrix is a square matrix that contains (column) inner products of some $M \times N$ matrix $\boldsymbol{A}$ such that
  $$\boldsymbol{G}_{\boldsymbol{A}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{H}}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_1^{\mathrm{H}}\boldsymbol{a}_N \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_M^{\mathrm{H}}\boldsymbol{a}_N & \cdots & \boldsymbol{a}_N^{\mathrm{H}}\boldsymbol{a}_N \end{pmatrix},$$
  where $\boldsymbol{v}_i$ is the ith column of $\boldsymbol{V}$.

- **Nonnegative matrix**
  A nonnegative matrix is a matrix whose values are nonnegative, i.e.,
  $$\boldsymbol{A} \geq 0 \iff a_{i,j} \geq 0 \; \forall i, j \in \mathbb{N}_{1:N}.$$

- **Permutation matrix**
  A permutation matrix is a square matrix with exactly a single one in each row and column with all the rest of the elements being zero. It can be obtained by permuting the columns of an identity matrix of the same dimensions.

- **Positive-definite matrix**
  A positive-definite matrix is a Hermitian matrix whose eigenvalues are positive.

- **Sparse matrix**
  A an $M \times N$ matrix $\boldsymbol{A}$ is called sparse if it has relatively few non-zero elements, i.e., $\sum_{n=1}^{N} \|\boldsymbol{a}_n\|_0 \ll MN$ where $\boldsymbol{a}_n$ is the nth column of $\boldsymbol{A}$.
- **Symmetric matrix**
  A symmetric matrix is a square matrix that is equal to its transpose, i.e., $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.
- **Unitary matrix**
  A unitary matrix is a matrix whose inverse is equal to its conjugate transpose, i.e., $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{H}}$.
- **Toeplitz matrix**
  A Toeplitz matrix is a matrix whose descending diagonals from left to right are constant. A particularly common class of Toeplitz matrices constitute square symmetric Toeplitz matrices that have a following structure

$$\boldsymbol{T} = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{N-1} \\ \alpha_1 & \alpha_0 & \alpha_1 & \cdots & \alpha_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{N-1} & \alpha_{N-2} & \alpha_{N-3} & \cdots & \alpha_0 \end{pmatrix}.$$

- **Vandermonde matrix**
  A Vandermonde matrix is a matrix whose rows contain the terms of a geometric progression of some variable α_i, i.e.,

$$\boldsymbol{V} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{N-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_M & \alpha_M^2 & \cdots & \alpha_M^{N-1} \end{pmatrix}.$$

Appendix D

Special functions and operations

Dirac delta function

The Dirac delta function denoted as $\delta(t)$ is a generalized function that is zero everywhere but at $t = 0$ that satisfies

$$\int_{-\infty}^{\infty} \delta(t)\mathsf{d}t = 1. \tag{D.1}$$

From (D.1) follows the *sifting* property of $\delta(t)$: for a non-zero scalar τ and some function $f(t)$

$$\int_{-\infty}^{\infty} \delta(t-\tau)f(t)\mathsf{d}t = f(\tau). \tag{D.2}$$

Gamma function

The gamma function is defined for all complex numbers except the non-positive integers as

$$\Gamma(t) \triangleq \int_0^{\infty} x^{t-1}e^{-x}\mathsf{d}t. \tag{D.3}$$

Note that if $n \in \mathbb{N}$, then $\Gamma(n) = (n-1)!$.

Polygamma function

The polygamma function of order m is defined as the $(m+1)$th derivative of the logarithm of the gamma function

$$\psi^{(m)}(z) \triangleq \frac{\mathsf{d}^{m+1}\ln(\Gamma(z))}{\mathsf{d}^{m+1}z}. \tag{D.4}$$

The polygamma functions of order $m = 0$ and $m = 1$ are often called the digamma and the trigamma functions, respectively.

Digamma function: The digamma function is commonly simply denoted as $\psi(z)$ and it can be calculated as

$$\psi^{(0)}(z) = \psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}. \tag{D.5}$$

Trigamma function: The trigamma function is also sometimes defined via the series expansion as

$$\psi^{(1)}(z) = \sum_{n=0}^{\infty} \frac{1}{(z+n)^2}. \tag{D.6}$$

Q-function

The Q-function $Q(x)$ is the compliment of the cumulative distribution function (CDF) of the standard normal distribution, i.e.,

$$Q(x) = 1 - \Phi(x) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \mathsf{d}t, \tag{D.7}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \mathsf{d}t$ is the CDF of $\mathcal{N}(0,1)$. For a standard normal random variable X, the Q-function gives the probability $\Pr[X > x]$.

Sign function

The sign function extracts the sign of a real number a such that

$$\operatorname{sign}(a) = \begin{cases} -1, & a < 0 \\ 0, & a = 0 \\ +1, & a > 0 \end{cases} \tag{D.8}$$

Modulo operation

For two real numbers a and b the modulo operation returns a remainder of the division of a by b. Formally, we define it as

$$a_{\operatorname{mod}_b} = |a| - \left| \left\lfloor \frac{a}{b} \right\rfloor b \right|, \tag{D.9}$$

where $\lfloor \cdot \rfloor$ is the operation of rounding to the nearest smaller integer.

Appendix E

Probability distributions

E.1 Continuous distributions

Normal Gaussian distribution

The normal Gaussian distribution is a continuous probability distribution with the probability density function (PDF)

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{E.1}$$

where μ is the expectation of x and σ^2 is its variance. When a random variable x is distributed according to (E.1) we express it by writing

$$x \sim \mathcal{N}\left(\mu,\sigma^2\right). \tag{E.2}$$

Finally, a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$ is commonly referred to as the standard normal Gaussian distribution and it is denoted as $\mathcal{N}(0,1)$.

Multivariate Gaussian distribution

The multivariate normal Gaussian distribution is a generalization of a (one-dimensional) normal Gaussian distribution to higher dimensions. A random vector $\boldsymbol{x} = [x_1,\ldots,x_N]^{\mathrm{T}}$ is said to be normally distributed if for any vector $\boldsymbol{a} \in \mathbb{R}^N$ the random variable $y = \boldsymbol{a}^{\mathrm{T}}$ is normally distributed. Multivariate normal distribution is parametrized by the mean vector

$$\boldsymbol{\mu} = \mathbb{E}\{\boldsymbol{x}\} = [\mathbb{E}\{x_1\},\ldots,\mathbb{E}\{x_N\}]^{\mathrm{T}}, \tag{E.3}$$

and the $N \times N$ covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}\left\{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right\}. \tag{E.4}$$

One denotes it as $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

When all the elements $x_n, n = 1, \ldots, N$ are i.i.d. normal random variables with the same mean and variance, i.e., $x_n \sim \mathcal{N}(\mu, \sigma)\ \forall n$, then we use the univariate notation $\boldsymbol{x} \sim \mathcal{N}(\mu, \sigma)$ to denote the distribution of $\boldsymbol{x}$ and say that $\boldsymbol{x}$ is normally distributed with the mean μ and variance σ^2.

Complex Gaussian distribution

The complex normal distribution characterizes complex random vectors $\boldsymbol{x} = \boldsymbol{a} + \jmath\boldsymbol{b} \in \mathbb{C}^N$ whose real and imaginary parts are jointly normal distributed. It is described by three parameters:

- the location parameter $\boldsymbol{\mu} = \mathbb{E}\{\boldsymbol{x}\} \in \mathbb{C}^N$;
- the covariance $\boldsymbol{\Gamma} = \mathbb{E}\{(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^*\} \in \mathbb{C}^{N\times N}$;
- the relation parameter $\boldsymbol{C} = \mathbb{E}\left\{(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\right\} \in \mathbb{C}^{N\times N}$.

The complex normal distribution is called circularly-symmetric complex normal if $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{C} = \boldsymbol{0}$ and it is denoted $\mathcal{CN}(0, \boldsymbol{\Gamma})$. Furthermore, when all $\boldsymbol{a}$ and $\boldsymbol{b}$ are i.i.d. normal with equal mean μ and variance $\sigma^2/2$ we denote it as $\boldsymbol{x} \sim \mathcal{CN}(\mu, \sigma)$.

Gamma distribution

The Gamma distribution is a two-parameter probability distribution with a PDF given by

$$f(x|k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, \tag{E.5}$$

where $k, \theta > 0$ are the so-called shape and scale parameters while $\Gamma(\cdot)$ denotes the Gamma function defined in D. A Gamma distributed random variable is denoted as $x \sim \Gamma(k, \theta)$ and its mean and variance are given by $\mathbb{E}\{x\} = k\theta$ and $\mathrm{var}\{x\} = k\theta^2$, respectively.

Chi-square distribution

The chi-square distribution with K degrees of freedom χ^2_K is the distribution of a sum of squares of K independent standard random variables. In other words, if $x_i \sim \mathcal{N}(0, 1)\ \forall i$ are i.i.d. random variables, then $y = \sum_{i=1}^{K} x_i^2 \sim \chi^2_K$. The PDF of the chi-square distribution is

$$f(y|K) = \begin{cases} \dfrac{y^{(K/2-1)} e^{-y/2}}{2^{K/2}\Gamma(K/2)} & , y > 0 \\ 0 & , \text{otherwise} \end{cases}, \tag{E.6}$$

where $\Gamma(\cdot)$ is the Gamma function defined in Appendix D. The mean and variance of $y \sim \chi^2_K$ are $\mathbb{E}\{y\} = K$ and $\mathrm{var}\{y\} = 2K$, respectively. Note that the chi-square distribution is a special case of the Gamma distribution described above.

Rayleigh distribution

Rayleigh distribution is a continuous probability distribution defined on the set of positive real numbers. Its PDF is given by

$$f(x|\sigma) = \frac{x^2}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}},\ x > 0, \tag{E.7}$$

where σ is the scaled parameter. A Rayleigh distribution is a distribution of $z = \sqrt{x^2 + y^2}$ when $x, y \sim \mathcal{N}\left(0, \sigma^2\right)$. The mean and variance of the Rayleigh distribution are

$$\mathbb{E}\left\{x\right\} = \sigma\sqrt{\frac{2}{\pi}},\ \text{and} \tag{E.8}$$

$$\text{var}\left\{x\right\} = \frac{4-\pi}{2}\sigma^2, \tag{E.9}$$

respectively.

Exponential-Gamma distribution

The exponential-gamma (EGa) distribution represents a continuous probability distribution that models $y = \log(x)$ whenever $x \sim \Gamma(\alpha, \beta)$. It is defined on the set of real numbers and parametrized by μ, $k > 0$ and $\theta > 0$, called a *location*, a *shape* and a *scale* parameters, respectively [338]. The PDF of EGa distribution is given by

$$f(x|\mu, \theta, k) = \frac{1}{\theta\Gamma(k)} e^{\left(-e^{\frac{x-\mu}{\theta}} + \frac{k(x-\mu)}{\theta}\right)}, \tag{E.10}$$

where $\Gamma(t)$ is the Gamma function defined in D. The mean of EGa is given by

$$\mathbb{E}\left\{x\right\} = \mu + \theta\psi(k), \tag{E.11}$$

where $\psi(k)$ is the digamma (polygamma function of order 0) function defined in D, whereas its variance can be calulcated as

$$\text{var}\left\{x\right\} = \theta^2\psi^{(1)}(k), \tag{E.12}$$

where $\psi(k)^{(1)}$ is the trigamma (polygamma function of order 1) function defined in D.

E.2 Discrete Distributions

Bernoulli distribution

Bernoulli distribution is a probability distribution of a (discrete) random variable x that takes the value 1 with probability p and he value 0 with probability $1-p$, i.e.,

$$\Pr[x=1] = p = 1 - \Pr[x=0].$$

The probability mass function (PMF) of the Bernoulli distribution is

$$f(x|p) = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{otherwise} \end{cases}. \tag{E.13}$$

When x is Bernoulli distributed with probability of 1 equal to p we denote it by $x \sim \mathcal{B}_e(p)$. Finally, the mean and variance of $x \sim \mathcal{B}_e(p)$ are given by $\mathbb{E}\{x\} = p$ and $\text{var}\{x\} = p(1-p)$.

Binomial distribution

The Binomial distribution $\mathcal{B}_i(N,p)$ is a distribution of a random variable y that is a sum of n i.i.d. Bernoulli distributed random variables, i.e., $y = \sum_{n=1}^{N} x_n : x_n \sim \mathcal{B}_e(p)$ for all $n \in [1, N]$. The PMF of the Binomial distribution is given by

$$f(y|N,p) = \Pr[y=n] = C_n^N p^n (1-p)^{N-n}, \tag{E.14}$$

where $C_N^n = \frac{N!}{n!(N-n)!}$ is the binomial coefficient, while $\mathbb{E}\{y\} = Np$ and $\text{var}\{y\} = Np(1-p)$. Finally, for any positive scalar $a > 0 \in \mathbb{R}$ a random variable $z = ay$ is distributed as scaled Binomial with mean $\mathbb{E}\{z\} = aNp$ and variance $\text{var}\{z\} = a^2 Np(1-p)$. We denote it as $z \sim \mathcal{B}_i^a(N,p)$.

Rademacher distribution

Rademacher distribution is a distribution of a random variable that takes a value ± 1 with 0.5 probability. The PMF of a Rademacher distribution is

$$f(x) = \begin{cases} 0.5, & x = 1 \\ 0.5, & x = -1 \\ 0, & \text{otherwise} \end{cases}, \tag{E.15}$$

while its mean and variance are given by 0 and 1, respectively.

List of abbreviations

A2I	analog-to-information
ADC	analog-to-digital converter
AIC	Akaike's information criterion
BOMP	block-OMP
BP	basis pursuit
BPDN	basis pursuit denoising
BPF	bandpass filter
CCDF	complimentary cumulative distribution function
CDF	cumulative distribution function
CFAR	constant false alarm rate
CLT	central limit theorem
CoSaMP	compressive sampling matching pursuit
CPD	canonical polyadic decomposition
CR	cognitive radio
CRB	Cramér-Rao bound
CS	compressed sensing
CTF	continuous-to-finite
DC	direct current
DCT	discrete cosine transform
DDR	direct decision rule
DF	direction finding
DFT	discrete Fourier transform
DoA	direction of arrival
DTFT	discrete time Fourier transform

EADF	effective aperture distribution function
EDC	efficient detection criterion
EDR	energy-based decision rule
EFT	exponential fitting test
EGa	exponential-gamma
EMD	earth mover's distance
ETF	equiangular tight frames
ETT	empirical threshold test
IDTFT	inverse discrete time Fourier transform
IHT	iterative hard thresholding
IMV	infinite measurement vector
ISNR	input SNR
IST	iterative soft thresholding
LASSO	least absolute shrinkage and selection operator
LNA	low noise amplifier
LoS	line of sight
LPF	low-pass filter
LS	least-squares
MCS	multi-coset sampler
MDL	minimum description length
MED	multiband energy detector
ML	maximum likelihood
MMV	multiple measurement vector
MOS	model order selection
MP	matching pursuit
MSE	mean squared error
MUSIC	multiple signal classification
MWC	modulated wideband converter
NLoS	non line of sight
NP-hard	non-deterministic polynomial-time hard
NPR	Neyman-Pearson rule

OMP	orthogonal matching pursuit
OSNR	output SNR
PDF	probability density function
PMF	probability mass function
PSD	power spectral density
RA-OMP	rank aware orthogonal matching pursuit
ReMBo	reduce MMV and boost
RF	radio frequency
RIP	restricted isometry property
RMSE	root mean squared error
ROC	receiver operating characteristics
RSNR	recovered SNR
SAGE	space-alternating generalized expectation-maximization
SI	shift-invariant
SIE	stationarity interval estimation
SMV	single measurement vector
SNR	signal-to-noise ratio
SOE	sparsity order estimation
SOMP	simultaneous orthogonal matching pursuit
SP	subspace pursuit
SRR	support recovery rate
SVD	singular-value decomposition
TDoA	time difference of arrival
UCA	uniform circular array
ULA	uniform linear array
UWB	ultra wideband
WSS	wideband spectrum sensing
XOR	exclusive-or

List of notations

a	A scalar
$\boldsymbol{a}$	A vector
$\boldsymbol{A}$	A matrix
$\boldsymbol{\mathcal{A}}$	A tensor
$\mathbf{I}_M$	An $M \times M$ identity matrix
$\mathbf{1}_{M \times N}$	An $M \times M$ matrix of all ones
$\mathbf{1}_N$	A length-M vector of all ones
$\mathbf{0}_{M \times N}$	An $M \times M$ matrix of all zeros
$\mathbf{0}_N$	A length-M vector of all zeros
$\|\cdot\|_p$	An ℓ_p norm of a vector
$\|\cdot\|_{p,q}$	An $L_{p,q}$ norm of a matrix
$\|\cdot\|_{\mathrm{F}}$	A Frobenius norm of a matrix
$\lambda^*(\cdot)$	A Lebesgue measure of an interval
$\lfloor\cdot\rfloor$	Operation of rounding to the nearest lower integer
$\lceil\cdot\rceil$	Operation of rounding to the nearest upper integer
$[\cdot]$	Operation of rounding to the nearest integer
$\mathcal{O}(\cdot)$	On the order of
$\langle\cdot\rangle$	Vector inner product
$\odot$	Hadamard (element-wise) matrix product
$\diamond$	Khatri-Rao matrix product
$\otimes$	Kronecker matrix product
$\ominus$	Symmetric difference between two sets
$(\cdot)^*$	A vector or matrix complex conjugate

$(\cdot)^{\mathrm{T}}$	A vector or matrix transpose
$(\cdot)^{\mathrm{H}}$	A vector or matrix Hermitian transpose (transpose and complex conjugation)
$(\cdot)^{\dagger}$	The Moore-Penrose matrix pseudo-inverse
$\mathrm{rank}(\cdot)$	Rank of a matrix
$\mathrm{krank}(\cdot)$	Kruskal rank of a matrix
$\mathrm{spark}(\cdot)$	Spark of a matrix
$\mathrm{diag}(\cdot)$	A vector containing the main diagonal of a matrix
$\mathrm{trace}\,(\cdot)$	A sum of the diagonal elements of a square matrix
$\mathrm{vec}\,(\cdot)$	A vector containing the columns of a matrix stacked on top of one another
$\mathbb{C}$	A set of complex numbers
$\mathbb{Z}$	A set of integer numbers
$\mathbb{N}$	A set of natural numbers
$\mathbb{N}_0$	A set of natural numbers plus 0
$\mathbb{N}_{k:p}$	A set of natural numbers from $k \in \mathbb{N}$ to $p \in \mathbb{N}$
$\mathbb{R}$	A set of real numbers

Bibliography

Own publications as author and co-author

Journals

[A1] A. Lavrenko, F. Roemer, G. Del Galdo, and R. Thoma, „On the SNR variability in noisy compressed sensing“, *IEEE Signal Proc. Letters*, vol. 24, pp. 1148–1152, 2017.

[A2] A. Bollig, A. Lavrenko, M. Arts, and R. Mathar, „Compressive cyclostationary spectrum sensing with a constant false alarm rate“, *EURASIP Journal on Wireless Communications and Networking*, vol. 2017, no. 1, p. 13, 2017.

[A3] M. Ibrahim, V. Ramireddy, A. Lavrenko, J. König, F. Römer, M. Landmann, M. Grossmann, G. Del Galdo, and R. S. Thomä, „Design and analysis of compressive antenna arrays for direction of arrival estimation“, *Elsevier Signal Processing*, vol. 138, pp. 35–47, 2017.

[A4] R. K. Sharma, A. Lavrenko, D. Kolb, and R. S. Thomä, „Cognitive scout node for communication in disaster scenarios“, *Journal of Computer Networks and Communications*, vol. 2012, p. 11, 2012.

Conference proceedings

[A5] A. Lavrenko, F. Römer, G. Del Galdo, and R. Thomä, „Sensitivity analysis of compressed sensing matrix design to random perturbations“, in *26th European Signal Processing Conference (EUSIPCO 2018)*, Rome, Italy, Sep. 2018, to appear.

[A6] S. P. Prawar, A. Lavrenko, M. Ibrahim, F. Römer, G. Del Galdo, and R. S. Thomä, „Combining matrix design for 2D DoA estimation with compressive antenna arrays“, in *22nd International ITG Workshop on Smart Antennas (WSA 2018)*, Bochum, Germany, Mar. 2018, to appear.

[A7] R. Hincapie, C. Gomez, A. Lavrenko, L. Betancur, and J. Schmitz, „Sparse framework for hybrid TDOA/DOA multiple emitter localization“, in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2017)*, Bilbao, Spain, Dec. 2017, to appear.

[A8] A. Lavrenko, F. Römer, G. Del Galdo, and R. S. Thomä, „Multiband TDoA estimation from sub-Nyquist samples with distributed wideband sensing nodes“, in *IEEE Global Conference on Signal and Information Processing (GlobalSIP 2017)*, Montreal, Canada, Nov. 2017, pp. 96–100.

[A9] A. Lavrenko, F. Roemer, G. Del Galdo, and R. S. Thomä, „On the earth mover's distance as a performance metric for sparse support recovery“, in *IEEE Global Conference on Signal and Information Processing (GlobalSIP 2016)*, Greater Washington, D.C., USA, Dec. 2016, pp. 1368–1372.

[A10] A. Lavrenko, A. Sosa, A. Navarro, and R. S. Thomä, „Interference detection in centralized cooperative spectrum sensing from sub-Nyquist samples“, in *IEEE 27th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2016)*, Valencia, Spain, Sep. 2016, pp. 1366–1371.

[A11] A. Lavrenko, F. Roemer, S. Stein Ioushua, D. Cohen, G. Del Galdo, R. S. Thomä, and Y. C. Eldar, „Spatially resolved sub-Nyquist sensing of multiband signals with arbitrary antenna arrays“, in *IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2016)*, Edinburgh, UK, Jul. 2016, pp. 1–5.

[A12] A. Lavrenko, F. Römer, G. Del Galdo, and R. S. Thomä, „Sparsity order estimation for sub-Nyquist sampling and recovery of sparse multiband signals“, in *IEEE International Conference on Communications (ICC 2015)*, London, UK, Jun. 2015, pp. 4926–4931.

[A13] A. Lavrenko, A. Bollig, and R. S. Thomä, „Compressive energy detection for blind coarse wideband sensing: comparative performance study“, in *IEEE 12th International Symposium on Wireless Communication Systems (ISWCS 2015)*, Brussels, Belgium, Aug. 2015, pp. 491–495.

[A14] A. Lavrenko, F. Römer, G. Del Galdo, R. S. Thomä, and O. Arikan, „Detection of time-varying support via rank evolution approach for effective joint sparse recovery“, in *23rd European Signal Processing Conference (EUSIPCO 2015)*, Nice, France, Aug. 2015, pp. 1726–1730.

[A15] B. Özer, A. Lavrenko, S. Gezici, F. Römer, G. D. Galdo, and O. Arikan, „Adaptive measurement matrix design for compressed DoA estimation with sensor arrays“, in *Proceedings 49-th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2015, pp. 1769–1773.

[A16] A. Lavrenko, F. Römer, G. Del Galdo, R. S. Thomä, and O. Arikan, „An empirical eigenvalue-threshold test for sparsity level estimation from compressed measurements“, in *22nd European Signal Processing Conference (EUSIPCO 2014)*, Lisbon, Portugal, Sep. 2014, pp. 1761–1765.

[A17] A. Lavrenko, F. Römer, G. Del Galdo, and R. S. Thomä, „On the sensing matrix performance for support recovery of noisy sparse signals“, in *IEEE Global Conf. on Signal and Information Processing (GlobalSIP 2014)*, Atlanta, USA, 2014, pp. 679–683.

[A18] F. Römer, A. Lavrenko, G. D. Galdo, T. Hotz, O. Arikan, and R. S. Thomä, „Sparsity order estimation for single snapshot compressed sensing“, in *Proc. 48th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, Nov. 2014, pp. 1220–1224.

[A19] R. K. Sharma, A. Lavrenko, and R. S. Thoma, „Performance evaluation of energy detection for correlated random signals“, in *77th IEEE Vehicular Technology Conference (VTC Spring 2013)*, 2013, pp. 1–5.

[A20] P. dos Santos, M. A. A. Kalil, O. Artemenko, A. Lavrenko, and A. Mitschele-Thiel, „Self-organized common control channel design for cognitive radio ad hoc networks“, in *IEEE 24th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2013)*, London, United Kingdom, Sep. 2013, pp. 2434–2438.

References by other authors

[1] R. A. DeVore, „Nonlinear approximation“, *Acta numerica*, vol. 7, pp. 51–150, 1998.

[2] B. K. Natarajan, „Sparse approximate solutions to linear systems“, *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[3] A. Rosenthal, D. Razansky, and V. Ntziachristos, „Quantitative optoacoustic signal extraction using sparse signal representation“, *IEEE Transactions on Medical Imaging*, vol. 28, no. 12, pp. 1997–2006, 2009.

[4] M. Protter and M. Elad, „Image sequence denoising via sparse and redundant representations“, *IEEE Trans. on Image Processing*, vol. 18, no. 1, pp. 27–35, 2009.

[5] A. M. Bruckstein, D. L. Donoho, and M. Elad, „From sparse solutions of systems of equations to sparse modeling of signals and images“, *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.

[6] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge university press, 2010.

[7] M. Elad, „Sparse and redundant representation modeling — What next?“, *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 922–928, 2012.

[8] M. Nelson and J.-L. Gailly, *The data compression book*. M&t Books New York, 1996.

[9] S. Saha, „Image compression—from DCT to wavelets: a review“, *Crossroads*, vol. 6, no. 3, pp. 12–21, 2000.

[10] D. L. Donoho and M. Elad, „Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell - 1$ minimization“, *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[11] E. J. Candès, J. Romberg, and T. Tao, „Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information“, *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[12] E. J. Candès, „Compressive sampling“, in *Proceedings of the international congress of mathematicians*, Madrid, Spain, vol. 3, 2006, pp. 1433–1452.

[13] E. J. Candès, J. Romberg, and T. Tao, „Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information“, *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[14] E. J. Candes and T. Tao, „Near-optimal signal recovery from random projections: Universal encoding strategies?“, *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[15] R. G. Baraniuk, „Compressive sensing“, *IEEE Signal Proc. Mag.*, vol. 24, no. 4, 2007.

[16] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, „Compressed sensing MRI“, *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[17] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, „Single-pixel imaging via compressive sampling“, *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[18] E. Candes and M. Wakin, „An Introduction To Compressive Sampling“, *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[19] J. Mark and T. Todd, „A nonuniform sampling approach to data compression“, *IEEE Transactions on Communications*, vol. 29, no. 1, pp. 24–32, 1981.

[20] J. Gansman, J. Krogmeier, and M. Fitz, „Single frequency estimation with non-uniform sampling“, in *Thirtieth Asilomar Conference on Signals, Systems and Computers*, IEEE, vol. 1, 1996, pp. 399–403.

[21] N. Sayiner, H. V. Sorensen, and T. Viswanathan, „A non-uniform sampling technique for A/D conversion“, in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1993, pp. 1220–1223.

[22] P. Feng and Y. Bresler, „Spectrum-blind minimum-rate sampling and reconstruction of multiband signals“, in *IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 1996, pp. 1688–1691.

[23] R. Venkataramani and Y. Bresler, „Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals“, *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2301–2313, 2001.

[24] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge University Press, 2015.

[25] J. A. Pugh, R. J. C. Bultitude, and P. J. Vigneron, „Propagation measurements and modelling for multiband communications on tactical VHF channels“, in *MILCOM 2007 - IEEE Military Communications Conference*, 2007, pp. 1–7.

[26] C.-H. Yun, A.-R. Cho, S.-G. Kim, J.-W. Park, and Y.-K. Lim, „Design of multi-band maritime network for ships and its applications“, *Journal of Information and Communication Convergence Engineering*, vol. 7, no. 3, pp. 314–322, 2009.

[27] H. Song and W. Hodgkiss, „Efficient use of bandwidth for underwater acoustic communication“, *Journal of the Association Society of America*, vol. 134, no. 2, pp. 905–908, 2013.

[28] D. Cohen, A. Dikopoltsev, R. Ifraimov, and Y. C. Eldar, „Towards sub-Nyquist cognitive radar“, in *IEEE Radar Conference (RadarConf)*, 2016, pp. 1–4.

[29] M. Edrich and A. Schroeder, „Multiband multistatic passive radar system for airspace surveillance: A step towards mature PCL implementations“, in *International Conference on Radar*, IEEE, 2013, pp. 218–223.

[30] J. Balakrishnan, A. Batra, and A. Dabak, „A multi-band OFDM system for UWB communication“, in *IEEE Conference on Ultra Wideband Systems and Technologies*, 2003, pp. 354–358.

[31] G. R. Aiello and G. D. Rogerson, „Ultra-wideband wireless systems“, *IEEE Microwave Magazine*, vol. 4, no. 2, pp. 36–47, 2003.

[32] W. P. Siriwongpairat and K. R. Liu, *Ultra-wideband communications systems: multiband OFDM approach*. John Wiley & Sons, 2007.

[33] A. Batra, S. Lingam, and J. Balakrishnan, „Multi-band OFDM: a cognitive radio for UWB“, in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2006, p. 4.

[34] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge Univ. Press, 2012.

[35] Y. M. Lu and M. N. Do, „A theory for sampling signals from a union of subspaces“, *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2334–2345, 2008.

[36] Y. C. Eldar and M. Mishali, „Robust recovery of signals from a structured union of subspaces“, *IEEE Tran. on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[37] T. Blumensath and M. E. Davies, „Sampling theorems for signals from the union of finite-dimensional linear subspaces“, *IEEE Trans. on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.

[38] T. Blumensath, „Sampling and reconstructing signals from a union of linear subspaces“, *IEEE Trans. on Information Theory*, vol. 57, no. 7, pp. 4660–4671, 2011.

[39] M. Mishali, Y. C. Eldar, and A. J. Elron, „Xampling: Signal acquisition and processing in union of subspaces“, *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4719–4734, 2011.

[40] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[41] D. L. Donoho, „Compressed sensing“, *IEEE Transactions on Information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[42] M. F. Duarte and R. G. Baraniuk, „Spectral compressive sensing“, *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

[43] L. He and L. Carin, „Exploiting structure in wavelet-based Bayesian compressive sensing“, *IEEE Trans. on Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.

[44] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics, Berlin, 2001, vol. 1.

[45] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, „Model-based compressive sensing“, *IEEE Trans. on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[46] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, „Recovery of clustered sparse signals from compressive measurements“, DTIC Document, Tech. Rep., 2009.

[47] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, „Block-sparse signals: Uncertainty relations and efficient recovery“, *IEEE Trans. on Sign. Proc.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[48] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, „Distributed compressed sensing“, *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[49] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, „Sparse solutions to linear inverse problems with multiple measurement vectors“, *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[50] D. Mo and M. F. Duarte, „Compressive parameter estimation with earth mover's distance via K-median clustering“, in *SPIE Optical Engineering+ Applications*, International Society for Optics and Photonics, 2013, 88581P–88581P.

[51] D. Ramasamy, S. Venkateswaran, and U. Madhow, „Compressive parameter estimation in AWGN“, *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 2012–2027, 2014.

[52] D. Mo and M. F. Duarte, „Performance of Compressive Parameter Estimation via K-Median Clustering“, *ArXiv e-prints*, Dec. 2015.

[53] E. J. Candes and T. Tao, „Decoding by linear programming“, *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[54] M. A. Davenport, „Random observations on random observations: Sparse signal acquisition and processing“, PhD thesis, Citeseer, 2010.

[55] A. M. Tillmann and M. E. Pfetsch, „The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing“, *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, 2014.

[56] S. G. Mallat and Z. Zhang, „Matching pursuits with time-frequency dictionaries“, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[57] J. A. Tropp and A. C. Gilbert, „Signal recovery from random measurements via orthogonal matching pursuit“, *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[58] L. Welch, „Lower bounds on the maximum cross correlation of signals“, *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 397–399, 1974.

[59] T. Strohmer and R. W. Heath, „Grassmannian frames with applications to coding and communication“, *Applied Computational Harmonic Analysis*, vol. 14, no. 3, pp. 257–275, 2003.

[60] V. N. Malozemov and A. Pevnyi, „Equiangular tight frames“, *Journal of Mathematical Sciences*, vol. 157, no. 6, 2009.

[61] E. Candes and J. Romberg, „Sparsity and incoherence in compressive sampling“, *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[62] D. L. Donoho, M. Elad, and V. N. Temlyakov, „Stable recovery of sparse overcomplete representations in the presence of noise“, *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[63] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, „A simple proof of the restricted isometry property for random matrices“, *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[64] R. Vershynin, „Introduction to the non-asymptotic analysis of random matrices“, *Compressed sensing: theory and applications*, 2012.

[65] H. Rauhut, „Compressive sensing and structured random matrices“, *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 1–92, 2010.

[66] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, „Introduction to compressed sensing“, *Compressed sensing: theory and applications*, 2012.

[67] X. Feng and Z. Zhang, „The rank of a random matrix“, *Applied mathematics and computation*, vol. 185, no. 1, pp. 689–694, 2007.

[68] M. A. Herman and T. Strohmer, „High-resolution radar via compressed sensing“, *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[69] A. Cohen, W. Dahmen, and R. DeVore, „Compressed sensing and best k-term approximation“, *Journal of the American mathematical society*, vol. 22, no. 1, pp. 211–231, 2009.

[70] B. Alexeev, J. Cahill, and D. G. Mixon, „Full spark frames“, *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1167–1194, 2012.

[71] R. A. DeVore, „Deterministic constructions of compressed sensing matrices“, *Journal of Complexity*, vol. 23, no. 4, pp. 918–925, 2007.

[72] P. Indyk, „Explicit constructions for compressed sensing of sparse signals“, in *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics, 2008, pp. 30–33.

[73] J. Haupt, L. Applebaum, and R. Nowak, „On the restricted isometry of deterministically subsampled Fourier matrices“, in *44th Annual Conference on Information Sciences and Systems (CISS)*, IEEE, 2010, pp. 1–6.

[74] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, D. Kutzarova, *et al.*, „Explicit constructions of RIP matrices and related problems“, *Duke Mathematical Journal*, vol. 159, no. 1, pp. 145–185, 2011.

[75] L. Applebaum, S. D. Howard, S. Searle, and R. Calderbank, „Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery“, *Applied and Computational Harmonic Analysis*, vol. 26, no. 2, pp. 283–290, 2009.

[76] R. Calderbank, S. Howard, and S. Jafarpour, „Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property“, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 358–374, 2010.

[77] A. Amini and F. Marvasti, „Deterministic construction of binary, bipolar, and ternary compressed sensing matrices“, *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2360–2370, 2011.

[78] J. Haupt and R. Nowak, „Signal reconstruction from noisy random projections“, *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[79] Z. Ben-Haim, Y. C. Eldar, and M. Elad, „Coherence-based performance guarantees for estimating a sparse vector under random noise“, *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.

[80] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, „The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range“, *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4628–4642, 2012.

[81] S. Muthukrishnan *et al.*, „Data streams: Algorithms and applications“, *Foundations and Trends® in Theoretical Computer Science*, vol. 1, no. 2, pp. 117–236, 2005.

[82] D. L. Donoho and P. B. Stark, „Uncertainty principles and signal recovery“, *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[83] S. S. Chen, D. L. Donoho, and M. A. Saunders, „Atomic decomposition by basis pursuit“, *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[84] D. L. Donoho and X. Huo, „Uncertainty principles and ideal atomic decomposition“, *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[85] M. Rudelson and R. Vershynin, „Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements“, in *40th Annual Conference on Information Sciences and Systems*, IEEE, 2006, pp. 207–212.

[86] S. Levy and P. K. Fullagar, „Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution“, *Geophysics*, vol. 46, no. 9, pp. 1235–1243, 1981.

[87] E. J. Candes, „The restricted isometry property and its implications for compressed sensing“, *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[88] E. J. Candes, J. K. Romberg, and T. Tao, „Stable signal recovery from incomplete and inaccurate measurements“, *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[89] E. J. Candès, Y. Plan, *et al.*, „Near-ideal model selection by l1 minimization“, *The Annals of Statistics*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[90] E. Candes and T. Tao, „The Dantzig selector: Statistical estimation when p is much larger than n“, *The Annals of Statistics*, pp. 2313–2351, 2007.

[91] V. Koltchinskii *et al.*, „The Dantzig selector and sparsity oracle inequalities“, *Bernoulli*, vol. 15, no. 3, pp. 799–828, 2009.

[92] R. Tibshirani, „Regression shrinkage and selection via the lasso“, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[93] N. Meinshausen and B. Yu, „Lasso-type recovery of sparse representations for high dimensional data“, *The Annals of Statistics*, pp. 246–270, 2009.

[94] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, „Simultaneous analysis of Lasso and Dantzig selector“, *The Annals of Statistics*, pp. 1705–1732, 2009.

[95] J. A. Tropp and S. J. Wright, „Computational methods for sparse solution of linear inverse problems“, *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[96] E. Van Den Berg and M. P. Friedlander, „Probing the Pareto frontier for basis pursuit solutions“, *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[97] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, „A wide-angle view at iterated shrinkage algorithms“, in *Optical Engineering+ Applications*, International Society for Optics and Photonics, 2007, pp. 670 102–670 102.

[98] T. Blumensath, M. E. Davies, and G. Rilling, „Greedy algorithms for compressed sensing“, *Compressed sensing: theory and applications*, pp. 348–393, 2012.

[99] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, „Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition“, in *The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, IEEE, 1993, pp. 40–44.

[100] G. Davis, S. Mallat, and M. Avellaneda, „Adaptive greedy approximations“, *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[101] J. A. Tropp, „Greed is good: Algorithmic results for sparse approximation“, *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[102] T. T. Cai and L. Wang, „Orthogonal matching pursuit for sparse signal recovery with noise“, *IEEE Trans. on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[103] D. Needell and J. A. Tropp, „Cosamp: iterative signal recovery from incomplete and inaccurate samples“, *Communications of the ACM*, vol. 53, no. 12, pp. 93–100, 2010.

[104] W. Dai and O. Milenkovic, „Subspace pursuit for compressive sensing: Closing the gap between performance and complexity“, DTIC Document, Tech. Rep., 2008.

[105] S. Sardy, A. G. Bruce, and P. Tseng, „Block coordinate relaxation methods for nonparametric wavelet denoising“, *Journal of computational and graphical statistics*, vol. 9, no. 2, pp. 361–379, 2000.

[106] I. Daubechies, M. Defrise, and C. De Mol, „An iterative thresholding algorithm for linear inverse problems with a sparsity constraint“, *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[107] J.-L. Starck, M. Elad, and D. Donoho, „Redundant multiscale transforms and their application for morphological component separation“, *Advances in Imaging and Electron Physics*, vol. 132, pp. 287–348, 2004.

[108] T. Blumensath and M. E. Davies, „Iterative thresholding for sparse approximations“, *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[109] M. Fornasier and H. Rauhut, „Iterative thresholding algorithms“, *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.

[110] T. Blumensath and M. E. Davies, „Iterative hard thresholding for compressed sensing“, *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[111] A. Maleki and D. L. Donoho, „Optimally tuned iterative reconstruction algorithms for compressed sensing“, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 330–341, 2010.

[112] M. A. Davenport and M. B. Wakin, „Analysis of orthogonal matching pursuit using the restricted isometry property“, *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.

[113] D. Needell and R. Vershynin, „Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit“, *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[114] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, „Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit“, *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[115] Y. Chi, L. Scharf, A. Pezeshki, and A. R. Calderbank, „Sensitivity to basis mismatch in compressed sensing“, *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[116] T. Strohmer, „Measure what should be measured: progress and challenges in compressive sensing“, *IEEE Signal Proc. Letters*, vol. 19, no. 12, pp. 887–893, 2012.

[117] M. Hegland, „Adaptive sparse grids“, *Anziam Journal*, vol. 44, pp. 335–353, 2003.

[118] D. Malioutov, M. Çetin, and A. S. Willsky, „A sparse signal reconstruction perspective for source localization with sensor arrays“, *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[119] L. Hu, Z. Shi, J. Zhou, and Q. Fu, „Compressed sensing of complex sinusoids: An approach based on dictionary refinement“, *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3809–3822, 2012.

[120] D. Shutin and B. H. Fleury, „Sparse variational Bayesian SAGE algorithm with application to the estimation of multipath wireless channels“, *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3609–3623, 2011.

[121] D. Shutin, W. Wang, and T. Jost, „Incremental sparse Bayesian learning for parameter estimation of superimposed signals“, in *10th International Conference on Sampling Theory and Applications*, 2013, pp. 6–9.

[122] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, „Dictionary learning algorithms for sparse representation“, *Neural computation*, vol. 15, no. 2, pp. 349–396, 2003.

[123] M. Yaghoobi, T. Blumensath, and M. E. Davies, „Dictionary learning for sparse approximations with the majorization method“, *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2178–2191, 2009.

[124] S. Ravishankar and Y. Bresler, „Learning sparsifying transforms“, *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1072–1086, 2013.

[125] O. Teke, A. C. Gurbuz, and O. Arikan, „Perturbed orthogonal matching pursuit“, *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6220–6231, 2013.

[126] K. Fyhn, S. H. Jensen, and M. F. Duarte, „Compressive time delay estimation using interpolation“, in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 624–624.

[127] M. Ibrahim, F. Romer, R. Alieiev, G. Del Galdo, and R. S. Thoma, „On the estimation of grid offsets in CS-based direction-of-arrival estimation“, in *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6776–6780.

[128] K. Fyhn, F. Marco, S. Holdt, *et al.*, „Compressive parameter estimation for sparse translation-invariant signals using polar interpolation“, *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 870–881, 2015.

[129] B. Recht, M. Fazel, and P. A. Parrilo, „Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization“, *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[130] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, „The convex geometry of linear inverse problems“, *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[131] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, „Compressed sensing off the grid“, *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[132] E. F. Complete, „On gridless sparse methods for line spectral estimation from complete and incomplete data“, *IEEE Transactions on Signal Processing*, p. 1, 2015.

[133] J. Chen and X. Huo, „Theoretical results on sparse representations of multiple measurement vectors“, *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[134] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, „Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit“, *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[135] J. A. Tropp, „Algorithms for simultaneous sparse approximation. Part II: Convex relaxation“, *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[136] M. E. Davies and Y. C. Eldar, „Rank awareness in joint sparse recovery“, *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.

[137] M. Mishali and Y. Eldar, „Reduce and boost: Recovering arbitrary sets of jointly sparse vectors“, *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4692–4702, 2008.

[138] S. Patterson, Y. C. Eldar, and I. Keidar, „Distributed sparse signal recovery for sensor networks“, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 4494–4498.

[139] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, „Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms“, *Journal of Fourier analysis and Applications*, vol. 14, no. 5-6, pp. 655–687, 2008.

[140] R. Schmidt, „Multiple emitter location and signal parameter estimation“, *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[141] Y. Bresler, „Spectrum-blind sampling and compressive sensing for continuous-index signals“, in *Information Theory and Applications Workshop*, IEEE, 2008, pp. 547–554.

[142] M. Fornasier and H. Rauhut, „Recovery algorithms for vector-valued data with joint sparsity constraints“, *SIAM Journ. on Num. Anal.*, vol. 46, no. 2, pp. 577–613, 2008.

[143] Y. C. Eldar and H. Rauhut, „Average case analysis of multichannel sparse recovery using convex relaxation“, *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[144] K. Lee, Y. Bresler, and M. Junge, „Subspace methods for joint sparse recovery“, *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.

[145] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jing, „Greedy Algorithms for Joint Sparse Recovery.“, *IEEE Trans. Signal Processing*, vol. 62, no. 7, pp. 1694–1704, 2014.

[146] M. Mishali, Y. Eldar, O. Dounaevsky, and E. Shoshan, „Xampling: Analog to digital at sub-Nyquist rates“, *Circuits, Devices Systems, IET*, vol. 5, no. 1, pp. 8–20, 2011.

[147] M. Mishali and Y. C. Eldar, „Xampling: compressed sensing of analog signals“, *Compressed sensing: theory and applications*, 2012.

[148] C. de Boor, R. A. Devore, and A. Ron, „The Structure of Finitely Generated Shift-Invariant Spaces in L_2“, 1992.

[149] Y. C. Eldar, „Compressed sensing of analog signals in shift-invariant spaces“, *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2986–2997, 2009.

[150] J.-J. Fuchs, „Recovery of exact sparse representations in the presence of bounded noise“, *IEEE Trans. on Information Theory*, vol. 51, no. 10, pp. 3601–3608, 2005.

[151] T. T. Cai, L. Wang, and G. Xu, „New bounds for restricted isometry constants“, *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4388–4394, 2010.

[152] E. J. Candes and Y. Plan, „A probabilistic and RIPless theory of compressed sensing“, *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[153] D. L. Donoho, A. Maleki, and A. Montanari, „The noise-sensitivity phase transition in compressed sensing“, *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6920–6941, 2011.

[154] Y. Wu and S. Verdú, „Optimal phase transitions in compressed sensing“, *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6241–6263, 2012.

[155] E. Arias-Castro and Y. C. Eldar, „Noise folding in compressed sensing“, *Signal Processing Letters, IEEE*, vol. 18, no. 8, pp. 478–481, 2011.

[156] M. Elad, „Optimized projections for compressed sensing“, *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[157] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, „Toeplitz-structured compressed sensing matrices“, in *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*, IEEE, 2007, pp. 294–298.

[158] Z. Ben-Haim and Y. C. Eldar, „The Cramér-Rao bound for estimating a sparse parameter vector“, *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, 2010.

[159] M. J. Wainwright, „Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting“, *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[160] W. Wang, M. J. Wainwright, and K. Ramchandran, „Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices“, *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2967–2979, 2010.

[161] G. Tang and A. Nehorai, „Performance analysis for sparse support recovery“, *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[162] H.-J. Zepernick and A. Finger, *Pseudo random signal processing: theory and application*. John Wiley & Sons, 2013.

[163] M. A. Herman and T. Strohmer, „General deviants: An analysis of perturbations in compressed sensing“, *IEEE Journal of Selected topics in signal processing*, vol. 4, no. 2, pp. 342–349, 2010.

[164] E. Israeli, S. Tsiper, D. Cohen, E. Shoshan, R. Hilgendorf, A. Reysenson, and Y. C. Eldar, „Hardware calibration of the modulated wideband converter“, in *2014 IEEE Global Communications Conference (GLOBECOM)*, 2014, pp. 948–953.

[165] J. D. Haupt, R. G. Baraniuk, R. M. Castro, and R. D. Nowak, „Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements“, in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, IEEE, 2009, pp. 1551–1555.

[166] J. N. Laska and R. G. Baraniuk, „Regime change: bit-depth versus measurement-rate in compressive sensing“, *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3496–3505, 2012.

[167] J. A. Tropp *et al.*, „An introduction to matrix concentration inequalities“, *Foundations and Trends® in Machine Learning*, vol. 8, no. 1-2, pp. 1–230, 2015.

[168] M. Taboga, *Lectures on probability theory and mathematical statistics*. CreateSpace Independent Pub., 2012.

[169] E. C. Pinheiro and S. L. P. Ferrari, „A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data“, *Journal of Statistical Computation and Simulation*, vol. 86, no. 11, pp. 2241–2261, 2016.

[170] K. Skretting and K. Engan, „Recursive least squares dictionary learning algorithm“, *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2121–2130, 2010.

[171] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.

[172] B. Babadi, N. Kalouptsidis, and V. Tarokh, „Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling“, *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, 2009.

[173] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, „On the achievability of Cramer–Rao bound in noisy compressed sensing“, *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, 2012.

[174] E. Kreyszig, *Advanced engineering mathematics*. New York: Wiley and Sons, 2005.

[175] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, „Sensing matrix optimization for block-sparse decoding“, *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4300–4312, 2011.

[176] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, „On optimization of the measurement matrix for compressive sensing“, in *18th European Signal Processing Conference*, IEEE, 2010, pp. 427–431.

[177] N. Cleju, „Optimized projections for compressed sensing via rank-constrained nearest correlation matrix“, *Applied and Computational Harmonic Analysis*, vol. 36, no. 3, pp. 495–507, 2014.

[178] Z. Lin, C. Lu, and H. Li, „Optimized Projections for Compressed Sensing via Direct Mutual Coherence Minimization“, *arXiv preprint*, 2015.

[179] *Matlab optimization toolbox*, https://mathworks.com/products/optimization.html.

[180] E. Levina and P. Bickel, „The earth mover's distance is the mallows distance: Some insights from statistics“, in *Eighth IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 2001, pp. 251–256.

[181] Y. Rubner, C. Tomasi, and L. J. Guibas, „A metric for distributions with applications to image databases“, in *Sixth InternationalIEEE Conference on Computer Vision*, 1998, pp. 59–66.

[182] H. Ling and K. Okada, „An efficient earth mover's distance algorithm for robust histogram comparison“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, 2007.

[183] Y. Rubner, C. Tomasi, and L. J. Guibas, „The earth mover's distance as a metric for image retrieval“, *Intern. journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.

[184] P. Indyk and E. Price, „K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance“, in *Proceedings of the forty-third annual ACM symposium on Theory of computing*, 2011, pp. 627–636.

[185] L. Schmidt, C. Hegde, and P. Indyk, „The constrained earth mover distance model, with applications to compressive sensing“, in *10th International Conference on Sampling Theory and Appl.(SAMPTA)*, 2013.

[186] N. B. Karahanoglu and H. Erdogan, „A comparison of termination criteria for A-OMP“, in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1449–1453.

[187] S. Kallummil and S. Kalyani, „Tuning Free Orthogonal Matching Pursuit“, *arXiv preprint arXiv:1703.05080*, 2017.

[188] P. T. Boufounos, M. F. Duarte, and R. G. Baraniuk, „Sparse Signal Reconstruction from Noisy Compressive Measurements Using Cross Validation“, in *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP)*, Madison, WI, 2007, pp. 299–303.

[189] R. Ward, „Compressed sensing with cross validation“, *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5773–5782, 2009.

[190] M. Lopes, „Estimating Unknown Sparsity in Compressed Sensing.“, *ICML (3)*, vol. 28, pp. 217–225, 2013.

[191] P. Indyk, „Stable distributions, pseudorandom generators, embeddings and data stream computation“, in *41st Annual Symposium on Foundations of Computer Science*, IEEE, 2000, pp. 189–197.

[192] V. Bioglio, T. Bianchi, and E. Magli, „On the fly estimation of the sparsity degree in compressed sensing using sparse sensing matrices“, in *IEEE International Conference onAcoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 3801–3805.

[193] C. D. Austin, R. Moses, J. Ash, and E. Ertin, „On the Relation Between Sparse Reconstruction and Parameter Estimation With Model Order Selection“, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 560–570, 2010.

[194] M. Grcesb, J. K. Nielsen, *et al.*, „Joint direction-of-arrival and order estimation in compressed sensing using angles between subspaces“, in *IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 449–452.

[195] Y. Wang, Z. Tian, and C. Feng, „Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios“, *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2116–2125, 2012.

[196] S. K. Sharma, S. Chatzinotas, and B. Ottersten, „Compressive sparsity order estimation for wideband cognitive radio receiver“, *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4984–4996, 2014.

[197] Z. Zhang, F. Liu, R. Du, C. Huang, and J. Sheng, „Trace-based sparsity order estimation with sparsely sampled random matrices“, *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4790–4799, 2016.

[198] D. Angelosante, G. B. Giannakis, and E. Grossi, „Compressed sensing of time-varying signals“, in *Intern. Conf. on Digital Signal Processing*, IEEE, 2009, pp. 1–8.

[199] J. Ziniel, L. C. Potter, and P. Schniter, „Tracking and smoothing of time-varying sparse signals via approximate belief propagation“, in *44th Asilomar Conf. on Signals, Systems and Computers*, IEEE, 2010, pp. 808–812.

[200] M. S. Asif, A. Charles, J. Romberg, and C. Rozell, „Estimation and dynamic updating of time-varying signals with sparse variations“, in *Intern. Conf. on Acoustics, Speech and Signal Processing*, IEEE, 2011, pp. 3908–3911.

[201] S. Semper, F. Römer, T. Hotz, and G. DelGaldo, „Sparsity Order Estimation from a Single Compressed Observation Vector“, *ArXiv e-prints*, Jun. 2017. arXiv: 1706.05278.

[202] M. Mishali and Y. Eldar, „Blind multiband signal reconstruction: Compressed sensing for analog signals“, *IEEE Trans. on Signal Proc.*, vol. 57, no. 3, pp. 993–1009, 2009.

[203] M. Duarte and Y. Eldar, „Structured Compressed Sensing: From Theory to Applications“, *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.

[204] P. Stoica and Y. Selen, „Model-order selection: a review of information criterion rules“, *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[205] J. Da Costa, A. Thakre, F. Roemer, and M. Haardt, „Comparison of model order selection techniques for high-resolution parameter estimation algorithms“, in *Proc. 54th International Scientific Colloquium (IWK'09), Ilmenau, Germany*, 2009.

[206] T. Anderson, „Multivariate statistical analysis“, *Willey and Sons, New York*, 1984.

[207] J. Grouffaud, P. Larzabal, and H. Clergeot, „Some properties of ordered eigenvalues of a Wishart matrix: application in detection test and model order selection“, in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings*, vol. 5, 1996, pp. 2463–2466.

[208] H. Akaike, „A new look at the statistical model identification“, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[209] J. Rissanen, „A universal prior for integers and estimation by minimum description length“, *The Annals of statistics*, pp. 416–431, 1983.

[210] M. Wax and T. Kailath, „Detection of signals by information theoretic criteria“, *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[211] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, „On detection of the number of signals in presence of white noise“, *Journ. of multiv. analysis*, vol. 20, no. 1, pp. 1–25, 1986.

[212] T. W. Anderson, „Asymptotic theory for principal component analysis“, *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, 1963.

[213] S. Kritchman and B. Nadler, „Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory“, *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3930–3941, 2009.

[214] W. Chen, K. M. Wong, and J. P. Reilly, „Detection of the number of signals: A predicted eigen-threshold approach“, *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1088–1098, 1991.

[215] A. Quinlan, J.-P. Barbot, P. Larzabal, and M. Haardt, „Model order selection for short data: An exponential fitting test (EFT)“, *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2006.

[216] J. P. C. L. da Costa, M. Haardt, F. Romer, and G. Del Galdo, „Enhanced model order estimation using higher-order arrays“, in *Proc. of the Forty-First Asilomar Conference on Signals, Systems and Computers*, IEEE, 2007, pp. 412–416.

[217] S. M. Kay, *Fundamentals of statistical signal processing, volume II: detection theory.* Prentice Hall, 1998.

[218] R. R. Nadakuditi and J. W. Silverstein, „Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples“, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 468–480, 2010.

[219] R. R. Nadakuditi and A. Edelman, „Sample eigenvalue based detection of high dimensional signals in white noise using relatively few samples“, *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2625–2638, 2008.

[220] J.-J. Fuchs, „Sparsity and uniqueness for some specific under-determined linear systems“, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 2005, pp. v–729.

[221] M. Dominguez-Jimenez, N. González-Prelcic, G. Vazquez-Vilar, and R. López-Valcarce, „Design of universal multicoset sampling patterns for compressed sensing of multiband sparse signals“, in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3337–3340.

[222] K. Konstantinides and K. Yao, „Statistical Analysis of Effective Singular Values in Matrix Rank Determination“, *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 5, pp. 757–763, May 1988.

[223] T. J. Shan, M. Wax, and T. Kailath, „On spatial smoothing for estimation of coherent signals“, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 806–811, 1985.

[224] M. F. Duarte and R. G. Baraniuk, „Kronecker Compressive Sensing“, *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, Feb. 2012.

[225] B. Alexeev, J. Cahill, and D. G. Mixon, „Full spark frames“, *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1167–1194, 2012.

[226] T. G. Kolda and B. W. Bader, „Tensor Decompositions and Applications“, *SIAM Review*, vol. 51, no. 3, pp. 455–500, Sep. 2009.

[227] T. G. Kolda and B. W. Bader, „Tensor decompositions and applications“, *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[228] F. Roemer, „Advanced Algebraic Concepts for Efficient Multi-Channel Signal Processing“, PhD thesis, Ilmenau University of Technology, Oct. 2012.

[229] M. Haardt, F. Roemer, and G. D. Galdo, „Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems“, *IEEE Trans. Sig. Proc.*, vol. 56, pp. 3198–3213, 7 Jul. 2008.

[230] J. P. C. L. D. Costa, F. Roemer, M. Haardt, and R. T. de Sousa Jr., „Multi-Dimensional Model Order Selection“, *EURASIP J. on Adv. Sig. Proc.*, Jul. 2011.

[231] N. D. Sidiropoulos and A. Kyrillidis, „Multi-way compressed sensing for sparse low-rank tensors“, *IEEE Sig. Proc. Letters*, vol. 19, no. 11, pp. 757–760, 2012.

[232] C. F. Caiafa and A. Cichocki, „Multidimensional compressed sensing and their applications“, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 6, pp. 355–380, Nov. 2013.

[233] G. Yuan, X. Zhang, W. Wang, and Y. Yang, „Carrier aggregation for LTE-advanced mobile communication systems“, *IEEE Commun. Magazine*, vol. 48, no. 2, 2010.

[234] K. I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L. G. U. Garcia, and Y. Wang, „Carrier aggregation for LTE-advanced: functionality and performance aspects“, *IEEE Communications Magazine*, vol. 49, no. 6, 2011.

[235] J. Mitola and G. Q. Maguire, „Cognitive radio: making software radios more personal“, *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[236] S. Haykin, „Cognitive radio: brain-empowered wireless communications“, *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[237] S. M. Mishra, R. Tandra, and A. Sahai, „The case for multiband sensing“, in *Proc. of the 54th Annual Allerton Conf. on Communication, Control, and Computing*, 2007.

[238] G. Hattab and M. Ibnkahla, „Multiband Spectrum Access: Great promises for future cognitive radio networks“, *Proc. of the IEEE*, vol. 102, no. 3, pp. 282–306, 2014.

[239] Z. Quan, S. Cui, H. V. Poor, and A. H. Sayed, „Collaborative wideband sensing for cognitive radios“, *IEEE Signal Processing Magazine*, vol. 25, no. 6, 2008.

[240] H. Venkataraman and G.-M. Muntean, *Cognitive radio and its application for next generation cellular and wireless networks*. Springer, 2012.

[241] J. Wang, M. Ghosh, and K. Challapali, „Emerging cognitive radio applications: A survey“, *IEEE Communications Magazine*, vol. 49, no. 3, 2011.

[242] M. H. Rehmani, A. C. Viana, H. Khalife, and S. Fdida, „A cognitive radio based internet access framework for disaster response network deployment“, in *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, IEEE, 2010, pp. 1–5.

[243] S. Mirabbasi and K. Martin, „Classical and modern receiver architectures“, *Communications Magazine, IEEE*, vol. 38, no. 11, pp. 132–139, 2000.

[244] D. Cabric and R. W. Brodersen, „Physical layer design issues unique to cognitive radio systems“, in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, 2005, pp. 759–763.

[245] V. Blaschke, T. Renk, and F. K. Jondral, „A cognitive radio receiver supporting wide-band sensing“, in *Communications, IEEE International Conference on*, IEEE, 2008, pp. 499–503.

[246] J. Brendel, S. Riess, A. Stoeckle, R. Rummel, and G. Fischer, „A spectrum sensing network for cognitive PMSE systems“, *Frequenz*, vol. 66, no. 9-10, pp. 269–277, 2012.

[247] R. Mzyk, G. Dehm-Andone, F. Hausknecht, G. Fischer, R. Weigel, and A. Koelpin, „Design aspects of scanning receiver systems“, in *The 7th German Microwave Conference (GeMiC)*, IEEE, 2012, pp. 1–4.

[248] R. Caromi, Y. Xin, and L. Lai, „Fast multiband spectrum scanning for cognitive radio systems“, *IEEE Trans.on Communications*, vol. 61, no. 1, pp. 63–75, 2013.

[249] B. Farhang-Boroujeny, „Filter bank spectrum sensing for cognitive radios“, *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1801–1811, 2008.

[250] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, „Wideband spectrum sensing in cognitive radio networks“, in *IEEE International Conference on Communications*, 2008, pp. 901–906.

[251] Y. Pei, Y.-C. Liang, K. C. Teh, and K. H. Li, „How much time is needed for wideband spectrum sensing?“, *IEEE Trans. on Wireless Communications*, vol. 8, no. 11, 2009.

[252] M. Lin, A. Vinod, and C. M. S. See, „A new flexible filter bank for low complexity spectrum sensing in cognitive radios“, *Journal of Signal Processing Systems*, vol. 62, no. 2, pp. 205–215, 2011.

[253] K. Smitha and A. Vinod, „A multi-resolution fast filter bank for spectrum sensing in military radio receivers“, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 7, pp. 1323–1327, 2012.

[254] M. Mishali and Y. C. Eldar, „Sub-Nyquist sampling“, *Signal Processing Magazine, IEEE*, vol. 28, no. 6, pp. 98–124, 2011.

[255] M. Mishali and Y. C. Eldar, „From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals“, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, 2010.

[256] D. D. Arianandа and G. Leus, „Compressive wideband power spectrum estimation“, *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4775–4789, 2012.

[257] D. Cohen and Y. C. Eldar, „Sub-Nyquist Sampling for power spectrum sensing in cognitive radios: A unified approach“, *IEEE Transactions on Signal Processing*, vol. 62, pp. 3897–3910, 2014.

[258] S. Kirolos, T. Ragheb, J. Laska, M. F. Duarte, Y. Massoud, and R. G. Baraniuk, „Practical issues in implementing analog-to-information converters“, in *International Workshop on System-on-Chip for Real-Time Applications*, IEEE, 2006, pp. 141–146.

[259] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, „Beyond Nyquist: Efficient sampling of sparse bandlimited signals“, *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[260] G. L. Fudge, R. E. Bland, M. A. Chivers, S. Ravindran, J. Haupt, and P. Pace, „A Nyquist folding analog-to-information receiver“, in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, IEEE, 2008, pp. 541–545.

[261] R. Maleh, G. L. Fudge, F. A. Boyle, and P. E. Pace, „Analog-to-information and the Nyquist folding receiver“, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 564–578, 2012.

[262] L. Melián-Gutiérrez, A. Garcia-Rodriguez, I. Pérez-Álvarez, and S. Zazo, „Compressive Narrowband Interference Detection for Wideband Cognitive HF Front-Ends“, *Wireless Personal Communications*, vol. 94, no. 3, pp. 1643–1660, 2017.

[263] Z. Tian and G. Giannakis, „A wavelet approach to wideband spectrum sensing for cognitive radios“, in *1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, 2006, pp. 1–5.

[264] B. Farhang-Boroujeny, „Filter Bank Spectrum Sensing for Cognitive Radios“, *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 1801–1811, 2008.

[265] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, „Optimal multiband joint detection for spectrum sensing in cognitive radio networks“, *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 1128–1140, 2009.

[266] R. López-Valcarce and G. Vazquez-Vilar, „Wideband spectrum sensing in cognitive radio: Joint estimation of noise variance and multiple signal levels“, in *IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, 2009, pp. 96–100.

[267] T.-H. Yu, S. Rodriguez-Parera, D. Markovic, and D. Cabric, „Cognitive radio wideband spectrum sensing using multitap windowing and power detection with threshold adaptation“, in *IEEE International Conference on Communications (ICC)*, 2010, pp. 1–6.

[268] S. Srinu, S. L. Sabat, and S. K. Udgata, „Wideband spectrum sensing based on energy detection for cognitive radio network“, in *World Congress on Information and Communication Technologies (WICT)*, IEEE, 2011, pp. 651–656.

[269] P. Paysarvi-Hoseini and N. C. Beaulieu, „Optimal wideband spectrum sensing framework for cognitive radio systems“, *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1170–1182, 2011.

[270] K. Hossain and B. Champagne, „Wideband spectrum sensing for cognitive radios with correlated subband occupancy“, *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 35–38, 2011.

[271] K. Hossain, A. Assra, and B. Champagne, „Multiband joint detection with correlated spectral occupancy in wideband cognitive radios“, in *International Symposium on Wireless Communication Systems (ISWCS)*, IEEE, 2012, pp. 541–545.

[272] M. Gautier, V. Berg, and D. Noguet, „Wideband frequency domain detection using Teager-Kaiser energy operator“, in *7th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, IEEE, 2012, pp. 332–337.

[273] S. K. Jayaweera, „Wideband Spectrum Sensing“, *Signal Processing for Cognitive Radios*, pp. 321–376, 2014.

[274] Z. Tian and G. B. Giannakis, „Compressed sensing for wideband cognitive radios“, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, 2007, pp. IV–1357.

[275] Y. Wang, Z. Tian, and C. Feng, „A two-step compressed spectrum sensing scheme for wideband cognitive radios“, in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, IEEE, 2010, pp. 1–5.

[276] M. Mishali and Y. C. Eldar, „Wideband spectrum sensing at sub-Nyquist rates [appl. corner]“, *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 102–135, 2011.

[277] D. Sundman, S. Chatterjee, and M. Skoglund, „On the use of compressive sampling for wide-band spectrum sensing“, in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2010, pp. 354–359.

[278] C.-P. Yen, Y. Tsai, and X. Wang, „Wideband spectrum sensing based on sub-Nyquist sampling“, *IEEE Trans. on Signal Processing*, vol. 61, no. 12, pp. 3028–3040, 2013.

[279] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, „Spectrum sensing for cognitive radio: State-of-the-art and recent advances“, *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[280] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, „Wideband spectrum sensing for cognitive radio networks: a survey“, *Wireless Communications, IEEE*, vol. 20, no. 2, pp. 74–81, 2013.

[281] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, „Optimal multiband joint detection for spectrum sensing in cognitive radio networks“, *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, 2009.

[282] M. A. Lexa, M. E. Davies, and J. S. Thompson, „Compressive and noncompressive power spectral density estimation from periodic nonuniform samples“, *arXiv preprint arXiv:1110.2722*, 2011.

[283] J. G. Proakis and D. G. Manolakis, *Digital Signal Proc.* Pearson Prentice Hall, 2007.

[284] S. Haykin, D. J. Thomson, and J. H. Reed, „Spectrum sensing for cognitive radio“, *Proceedings of the IEEE*, vol. 97, no. 5, pp. 849–877, 2009.

[285] T. Yucek and H. Arslan, „A survey of spectrum sensing algorithms for cognitive radio applications“, *Comm. Surveys & Tutorials, IEEE*, vol. 11, no. 1, pp. 116–130, 2009.

[286] H. Urkowitz, „Energy detection of unknown deterministic signals“, *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[287] *Xampling hardware*, http://webee.technion.ac.il/Sites/People/YoninaEldar/hardware.php.

[288] M. Mishali and Y. C. Eldar, „Expected RIP: Conditioning of the modulated wideband converter“, in *IEEE Information Theory Workshop*, 2009, pp. 343–347.

[289] W. Davies, *Generation and properties of maximum length sequences*, 1966.

[290] C.-P. Yen, Y. Tsai, and X. Wang, „Wideband spectrum sensing based on sub-Nyquist sampling“, *IEEE Trans. on Signal Processing*, vol. 61, no. 12, pp. 3028–3040, 2013.

[291] M. Alam, K. Jamil, Z. O. Alhekail, and S. Al-Humaidi, „A multi-band multi-beam software-defined passive radar. Part II: Signal processing“, 2012.

[292] P. P. Vaidyanathan and P. Pal, „Sparse Sensing With Co-Prime Samplers and Arrays“, *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[293] P. Pal and P. P. Vaidyanathan, „Nested Arrays: A Novel Approach to Array Processing With Enhanced Degrees of Freedom“, *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[294] K. Han, Y. Wang, B. Kou, and W. Hong, „Parameters estimation using a random linear array and compressed sensing“, in *International Congress on Image and Signal Processing (CISP)*, vol. 8, Oct. 2010, pp. 3950–3954.

[295] M. Rossi, A. M. Haimovich, and Y. C. Eldar, „Spatial compressive sensing in MIMO radar with random arrays“, in *Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2012, pp. 1–6.

[296] S. Shakeri, D. D. Arianandа, and G. Leus, „Direction of arrival estimation using sparse ruler array design“, in *International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2012, pp. 525–529.

[297] M. B. Hawes and W. Liu, „Compressive sensing-based approach to the design of linear robust sparse antenna arrays with physical size constraint“, *IET Microwaves, Antennas Propagation*, vol. 8, no. 10, pp. 736–746, Jul. 2014.

[298] Y. Wang, G. Leus, and A. Pandharipande, „Direction estimation using compressive sampling array processing“, in *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, IEEE, 2009, pp. 626–629.

[299] J. F. Gu, W. P. Zhu, and M. N. S. Swamy, „Compressed sensing for DOA estimation with fewer receivers than sensors“, in *IEEE International Symposium of Circuits and Systems (ISCAS)*, May 2011, pp. 1752–1755.

[300] D. D. Arianandа and G. Leus, „Compressive joint angular-frequency power spectrum estimation“, in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013, pp. 1–5.

[301] A. A. Kumar, S. G. Razul, and C.-M. S. See, „An efficient sub-Nyquist receiver architecture for spectrum blind reconstruction and direction of arrival estimation“, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2014, pp. 6781–6785.

[302] A. A. Kumar, S. G. Razul, and C.-M. S. See, „Carrier frequency and direction of arrival estimation with nested sub-Nyquist sensor array receiver“, in *23rd European Signal Processing Conference (EUSIPCO)*, IEEE, 2015, pp. 1167–1171.

[303] S. S. Ioushua, O. Yair, D. Cohen, and Y. C. Eldar, „CaSCADE: Compressed carrier and DOA estimation“, *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2645–2658, 2017.

[304] C. Cui, W. Wu, and W.-Q. Wang, „Carrier Frequency and DOA Estimation of Sub-Nyquist Sampling Multi-band Sensor Signals“, *IEEE Sensors Journal*, 2017.

[305] L. C. Godara and A. Cantoni, „Uniqueness and linear independence of steering vectors in array space“, *The Journal of the Acoustical Society of America*, vol. 70, no. 2, pp. 467–475, 1981.

[306] H. Gazzah and K. Abed-Meraim, „Optimum ambiguity-free directional and omnidirectional planar antenna arrays for DOA estimation“, *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3942–3953, 2009.

[307] X. Guo, S. Miron, D. Brie, S. Zhu, and X. Liao, „A CANDECOMP/PARAFAC perspective on uniqueness of DOA estimation using a vector sensor array“, *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3475–3481, 2011.

[314] M. Landmann and G. Del Galdo, „Efficient antenna description for MIMO channel modelling and estimation“, in *7th European Conference on Wireless Technology*, IEEE, 2004, pp. 217–220.

[308] S. Stein, O. Yair, D. Cohen, and Y. C. Eldar, „Joint spectrum sensing and direction of arrival recovery from sub-Nyquist samples“, in *16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 331–335.

[309] J. B. Kruskal, „Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics“, *Linear algebra and its applications*, vol. 18, no. 2, pp. 95–138, 1977.

[310] L.-H. Lim and P. Comon, „Multiarray signal processing: Tensor decomposition meets compressed sensing“, *Comptes Rendus Mecanique*, vol. 338, no. 6, pp. 311–320, 2010.

[311] N. D. Sidiropoulos and A. Kyrillidis, „Multi-way compressed sensing for sparse low-rank tensors“, *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 757–760, 2012.

[312] S. Friedland, Q. Li, and D. Schonfeld, „Compressive sensing of sparse tensors“, *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4438–4447, 2014.

[313] M. Landmann, W. Kotterman, and R. Thomä, „On the influence of incomplete data models on estimated angular distributions in channel characterisation“, *Measurements*, vol. 3, no. 1, p. 4, 2007.

[315] S. Yarkan and H. Arslan, „Exploiting location awareness toward improved wireless system design in cognitive radio“, *IEEE Comm. Magazine*, vol. 46, no. 1, 2008.

[316] H. Celebi and H. Arslan, „Cognitive positioning systems“, *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, 2007.

[317] H. Celebi and H. Arslan, „Enabling location and environment awareness in cognitive radios“, *Computer Communications*, vol. 31, no. 6, pp. 1114–1125, 2008.

[318] F. Kocak, H. Celebi, S. Gezici, K. A. Qaraqe, H. Arslan, and H. V. Poor, „Time-delay estimation in dispersed spectrum cognitive radio systems“, *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 675 959, 2010.

[319] R. R. Thomas, B. Zayen, R. Knopp, and B. T. Maharaj, „Multiband time-of-arrival positioning technique for cognitive radio systems“, in *22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, IEEE, 2011, pp. 2315–2319.

[320] M. Pourkhaatoun and S. A. Zekavat, „Concatenated spectrum multi-band TOA estimation“, in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, IEEE, 2011, pp. 1192–1196.

[321] K. K. Mada, H.-C. Wu, and S. S. Iyengar, „Efficient and robust EM algorithm for multiple wideband source localization“, *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 3071–3075, 2009.

[322] F. Gustafsson and F. Gunnarsson, „Positioning using time-difference of arrival measurements“, in *Acoustics, Speech, and Signal Processing. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, IEEE, vol. 6, 2003, pp. VI–553.

[323] J. Schmitz, F. Schröder, and R. Mathar, „TDOA fingerprinting for localization in non-line-of-sight and multipath environments“, in *International Symposium on Antennas and Propagation (ISAP 2015)*, 2015, pp. 1–4.

[324] J. Scheuing and B. Yang, „Disambiguation of TDoA estimates in multi-path multi-source environments (DATEMM)“, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, 2006.

[325] Z. Ren, Y. Huang, Q. Chen, and H. Li, „Modeling and simulation of fading, pathloss, and shadowing in wireless networks“, in *IEEE International Conference on Communications Technology and Applications*, Oct. 2009, pp. 335–343.

[326] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[327] J. M. Duarte-Carvajalino and G. Sapiro, „Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization“, *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, 2009.

[328] W. Chen, M. R. Rodrigues, and I. J. Wassell, „On the design of optimized projections for sensing sparse signals in overcomplete dictionaries“, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 3457–3460.

[329] C.-Y. Lu and D.-S. Huang, „Optimized projections for sparse representation based classification“, *Neurocomputing*, vol. 113, pp. 213–219, 2013.

[330] D. D. Arianandа, D. Romero, and G. Leus, „Cooperative compressive power spectrum estimation“, in *IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2014, pp. 97–100.

[331] D. Cohen, A. Akiva, B. Avraham, and Y. C. Eldar, „Centralized cooperative spectrum sensing from sub-Nyquist samples for cognitive radios“, in *IEEE International Conference on Communications (ICC)*, 2015, pp. 7486–7491.

[332] D. Cohen, A. Akiva, B. Avraham, S. Patterson, and Y. C. Eldar, „Distributed cooperative spectrum sensing from sub-Nyquist samples for cognitive radios“, in *IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 336–340.

[333] J. C. Chen, K. Yao, and R. E. Hudson, „Source localization and beamforming“, *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30–39, 2002.

[334] I. Olkin and A. W. Marshall, *Inequalities: theory of majorization and its applications*. Academic press, 2016, vol. 143.

[335] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to probability*. Athena Scientific Belmont, MA, 2002, vol. 1.

[336] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, „Parallel Factor Analysis in Sensor Array Processing“, *IEEE Trans. on Signal Proc.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[337] C. D. Meyer, *Matrix analysis and applied linear algebra*. Siam, 2000, vol. 2.

[338] *Wolfram Alpha - a computational knowledge engine*, https://reference.wolfram.com.